

GERMÁN CASTELLANOS DOMÍNGUEZ  
YURI SEMENOVICH SHINAKOV

# ANÁLISIS DE ALEATORIEDAD EN SEÑALES Y SISTEMAS

UNIVERSIDAD NACIONAL DE COLOMBIA  
SEDE MANIZALES

ISBN: 978-958-8280-13-4

© 2007 UNIVERSIDAD NACIONAL  
DE COLOMBIA SEDE MANIZALES

**AUTORES:**

**GERMÁN CASTELLANOS DOMÍNGUEZ**

Ingeniero en Radiocomunicaciones  
Ph.D en Radiotecnía  
Profesor Asociado  
Universidad Nacional de Colombia  
Sede Manizales

**YURI SEMENOVICH SHINAKOV**

Ingeniero en Radiotecnía  
Ph.D en Radiotecnía  
Doctor en Ciencias Técnicas  
Director Departamento de Sistemas Radiotécnicos  
Profesor Titular  
Universidad Técnica de Comunicaciones e Informática  
de Moscú

**REVISADO:**

**JULIO CÉSAR GARCÍA ÁLVAREZ**

Ingeniero Electrónico  
Magíster en Ingeniería Electrónica y de Computadores  
Profesor Asistente  
Universidad Nacional de Colombia  
Sede Manizales

**FRANKLIN ALEXANDER SEPÚLVEDA SEPÚLVEDA**

Ingeniero Electrónico  
Magíster en Automatización Industrial  
Profesor Asistente  
Universidad Nacional de Colombia  
Sede Manizales

**IMPRESO:**

Centro de Publicaciones  
Universidad Nacional de Colombia  
Sede Manizales

Junio de 2007  
Primera edición

Este libro está hecho con ayuda de KOMA-Script y LATEX

## Prefacio

El libro que se presenta a continuación es el fruto de varios años de trabajo y de investigación. Los más recientes, dedicados a plasmar en las páginas los conocimientos que se tienen de los temas, de una manera cuidadosa, pedagógica, y moderna. Los años anteriores, a consolidar las experiencias en el trabajo diario de señales y sistemas con los conceptos teóricos, adquiridos y madurados con antelación. La sensación que uno tiene al leer el documento es la de saborear un buen vino tinto francés, probablemente un Burdeos, que al envejecer ha sabido pulirse al extremo de ser perfecto, y que al *descorcharlo* se desprenden los aromas y los bouquets que contiene. El profesor Germán Castellanos ha sabido conjugar el aparato matemático teórico con la experiencia adquirida a través de los numerosos proyectos de investigación con aplicaciones a señales de voz, electrocardiográficas, etc. Su formación de postgrado en uno de los centros más representativos del mundo en el análisis de señales se lo ha permitido. Este es un libro escrito con la cabeza fría de la escuela rusa, pero a un pulso caliente, como el talante latino, colombiano, boyacense incluso.

El autor escogió para escribir el tema de señales, y basó su exposición en buenas uvas: los sistemas ortogonales, las variables aleatorias, y los procesos estocásticos de Markov. Podemos decir que el *coupage* es excelente, y se augura un producto de excepción. Sin embargo, es el trato mimoso del enólogo lo que va a conferir al vino su calidez y su suntuosidad. Así, el tratamiento que el profesor Castellanos da a los temas es cariñoso y de una claridad expositiva envidiable. De cada tema, se escogen ejemplos selectos, que nos dan la nota de cata de cada capítulo. Se evoluciona desde ejemplos claros y básicos, a los más elaborados, y así aprendemos que la obra del profesor es una obra completa. En efecto, la parte más matemática correspondiente a las demostraciones de los teoremas se deja solo entrever, primero, por la complejidad misma, y segundo, para no perder el hilo de la exposición. Es ya en esos matices cuando vemos la calidad de la obra, que respira un aire fresco y un tratamiento moderno. Vemos ahí que el autor ha pensado en este libro también como un documento de trabajo del estudiante de postgrado, y es ahí donde hay que agradecer el esfuerzo de este verdadero enólogo de las señales.

El libro trata temas difíciles, como son los de ergodicidad y descomposición espectral, la convergencia y la continuidad de procesos aleatorios. En el capítulo 4 se muestran los procesos de Markov, con la perspectiva moderna de los Modelos Ocultos de Markov, que tantas aplicaciones están encontrando en todas las áreas de la ciencia aplicada. El profesor Castellanos sabe qué vino quiere hacer degustar, y dedica unas páginas interesantes al entrenamiento, la clasificación y comparación de modelos. En ese momento podemos disfrutar del *bouquet* exquisito del libro. Otro momento donde el libro desprende todos sus aromas es en los capítulos finales, donde el autor nos indica el objetivo último de esta obra: los métodos de detección y filtración de señales aleatorias, y de sistemas dinámicos.

Después de leer y releer la obra, disfrutando de la calidad de la exposición, no queda más que agradecer al profesor Castellanos su labor, y la de sus colaboradores de su grupo de trabajo académico. Y esperar que el enólogo nos sorprenda de nuevo con otro muy exquisito libro, que sin duda estudiantes e investigadores van a demandar muy prontamente.

---

Gerard Olivar, Manizales, 2007



# Introducción

Several years ago I reached the conclusion that the Theory of Probability should no longer be treated as adjunct to statistics or noise or any terminal topic, but should be included in the basic training of all engineers and physicists as a separate course

---

Papoulis

El interés por el estudio de las señales aleatorias<sup>1</sup> sigue siendo actual, aún cuando su aplicación asociada a la tecnología e ingeniería cuenta con varias décadas de desarrollo. Particular importancia tienen los métodos y técnicas de análisis en presencia de aleatoriedad en señales y sistemas, que estén relacionados con las tecnologías de proceso discreto, en las cuales el empleo de métodos efectivos matemáticos se ha hecho masivo, gracias a la aparición de dispositivos digitales de alto rendimiento.

El estudio del fenómeno de aleatoriedad en señales y sistemas está directamente relacionado con el análisis de las propiedades de interacción cuando se tiene algún grado de incertidumbre o información incompleta. La principal fuente de aleatoriedad está asociada, entre otros factores, con las perturbaciones de medida, la presencia de señales extrañas, y tal vez el más importante, la información útil que llevan las mismas señales. En este sentido, es cada vez mayor el interés por el empleo de métodos de proceso digital de señales, basados en matemática estadística, que se orientan principalmente a la solución de dos tareas: la representación adecuada del conjunto de señales para la solución de un problema concreto, que tenga en cuenta las condiciones reales de la aplicación y, segundo, el proceso óptimo de las señales recogidas.

El material dispuesto en el presente texto describe las formas básicas de representación y proceso de señales aleatorias, con especial énfasis en los modelos de análisis matemático estadístico, que se considera, tienen un aporte significativo en la formulación y solución de aplicaciones en las áreas de sismología, análisis de bioseñales, sistemas de medida, sistemas de control y seguimiento, radiocomunicación, entre otros. Por eso, el contenido presentado corresponde a la evolución en el análisis de aleatoriedad desde las señales hasta su asociación con los respectivos sistemas de proceso. En cada sección se presentan, tanto ejemplos como problemas (incluyendo algunos ejercicios en el computador), que tienen como intención mejorar y complementar la percepción del material teórico analizado.

El contenido del texto es el siguiente: el capítulo I describe los fundamentos de representación determinística de señales y sistemas. El capítulo II presenta la caracterización de variables aleatorias y las técnicas básicas de estimación. Aunque el material de ambos capítulos se supone conocido, por ejemplo de los cursos de *Teoría de Señales* y *Teoría de Probabilidades*, es pertinente su inclusión a fin de recordar aspectos y definiciones importantes que se hacen necesarios para el entendimiento de los capítulos posteriores. El capítulo III describe las particularidades en la representación de procesos aleatorios en el tiempo, incluyendo su análisis experimental. Debido a la importancia de los procesos estocásticos de Markov, estos se analizan por a parte en el capítulo IV, con especial atención en el análisis experimental mediante los modelos ocultos de Markov. El capítulo V corresponde al análisis de la transformación de señales aleatorias en sistemas, así como la formación de procesos aleatorios mediante sistemas lineales. El capítulo VI presenta los principales métodos de detección y filtración de señales aleatorias, entendida como la optimización de la forma de representación de los dispositivos de proceso.

Por último, la necesidad de aumentar la efectividad del funcionamiento de los sistemas de control, en condiciones de factores aleatorios, ha estimulado el desarrollo de métodos de optimización orientada a

---

<sup>1</sup> *Aleatorio*. Del Latín *aleatorius*, propio del juego de dados. Real Academia de la Lengua

mejorar el registro y precisión de la información sobre las propiedades de la planta a diseñar, así como de los regímenes de su funcionamiento. En este sentido, el capítulo VII describe el análisis de sistemas estocásticos, en particular, se analiza la filtración Kalman.

El material está orientado a los estudiantes de posgrado, que requieran profundizar en el análisis de procesos aleatorios. Además, se puede emplear como material de ayuda en el curso de *Procesos Estocásticos*. El material teórico presentado tiene carácter de referencia, y por esto no se da la deducción de varias de las expresiones, brindándose la literatura necesaria para la profundización de cada tema en particular.

Finalmente, se agradece a los profesores Julio F. Suárez, Gerard Olivar y Sergey Adzhemov por sus consejos y correcciones hechos para dar mayor precisión y entendimiento del presente trabajo. Se aprecia también la colaboración que prestaron los estudiantes del programa posgrado, Jorge Jaramillo, Luis G. Sánchez, Julián D. Arias, Mauricio Álvarez, David Avendaño, Andrés F. Quiceno y todos aquellos, quienes pacientemente leyeron, siguieron y colaboraron con la evolución de los diferentes borradores del texto.

---

*Los Autores, 2007*

# Índice general

<b>1. Señales y sistemas</b>	<b>1</b>
1.1. Representación discreta de señales . . . . .	1
1.1.1. Espacio de representación de señales . . . . .	1
1.1.2. Descomposición en funciones ortogonales . . . . .	3
1.1.3. Ejemplos de conjuntos ortogonales completos . . . . .	3
1.2. Representación integral de señales y sistemas . . . . .	7
1.2.1. Densidad espectral . . . . .	7
1.2.2. Transformaciones lineales . . . . .	11
1.2.3. Representación integral de sistemas lineales . . . . .	13
1.2.4. Representación de sistemas no lineales . . . . .	18
1.3. Discretización de señales y sistemas . . . . .	25
1.3.1. Discretización uniforme . . . . .	25
1.3.2. Transformadas ortogonales discretas . . . . .	26
1.3.3. Representación ortogonal de sistemas en tiempo discreto . . . . .	27
1.4. Representación dinámica de sistemas . . . . .	31
1.4.1. Sistemas lineales en variables de estado . . . . .	31
1.4.2. Solución de la ecuación de estado . . . . .	37
<b>2. Variables aleatorias</b>	<b>41</b>
2.1. Valores y funciones de probabilidad . . . . .	41
2.1.1. Espacio probabilístico de variables . . . . .	41
2.1.2. Valores medios y momentos . . . . .	46
2.1.3. Funciones de probabilidad . . . . .	53
2.1.4. Modelos de funciones de probabilidad con dimensión múltiple . . . . .	67
2.1.5. Medidas de información en variables aleatorias . . . . .	78
2.2. Estimación en variables aleatorias . . . . .	85
2.2.1. Estimación puntual de momentos . . . . .	86
2.2.2. Intervalos de confianza . . . . .	91
2.2.3. Estimación de parámetros en la distribución normal . . . . .	94
2.2.4. Prueba de hipótesis . . . . .	96
2.2.5. Estimación de dependencias funcionales . . . . .	98
2.3. Teoría de decisión estadística . . . . .	105
2.3.1. Definiciones básicas . . . . .	105
2.3.2. Decisión Bayesiana . . . . .	110
2.3.3. Decisión no Bayesiana . . . . .	112
<b>3. Señales aleatorias</b>	<b>115</b>
3.1. Señales aleatorias en el tiempo . . . . .	115

3.1.1.	Estacionariedad de las señales aleatorias . . . . .	116
3.1.2.	Ergodicidad de las señales aleatorias . . . . .	121
3.1.3.	Descomposición espectral de señales aleatorias . . . . .	127
3.1.4.	Densidad espectral de potencia . . . . .	130
3.1.5.	Convergencia y continuidad de procesos aleatorios . . . . .	138
3.2.	Análisis experimental de señales estacionarias . . . . .	143
3.2.1.	Estimación de momentos . . . . .	144
3.2.2.	Estimación de los coeficientes en la descomposición K-L . . . . .	147
3.2.3.	Estimación de la densidad espectral de potencia . . . . .	150
3.2.4.	Estimación de parámetros de regresión . . . . .	153
3.3.	Estimación espectral no paramétrica . . . . .	157
3.3.1.	Método de los periodogramas . . . . .	157
3.3.2.	Algoritmo de cálculo del método de periodograma . . . . .	164
3.3.3.	Ventanas de estimación espectral . . . . .	166
3.4.	Estimación espectral paramétrica . . . . .	169
3.4.1.	Reacción de un sistema lineal a una señal estacionaria . . . . .	169
3.4.2.	Modelos paramétricos de señales aleatorias . . . . .	170
3.4.3.	Estimación de los parámetros a partir de la función de correlación . . . . .	172
<b>4.</b>	<b>Procesos de Markov</b>	<b>177</b>
4.1.	Definición y clasificación . . . . .	177
4.1.1.	Cadenas de Markov . . . . .	178
4.1.2.	Procesos discretos de Markov . . . . .	183
4.1.3.	Procesos continuos de Markov . . . . .	187
4.1.4.	Sucesiones de Markov . . . . .	198
4.2.	Análisis experimental de procesos de Markov . . . . .	201
4.2.1.	Modelos ocultos de Markov . . . . .	201
4.2.2.	Entrenamiento del modelo . . . . .	203
4.2.3.	Clasificación y comparación modelos . . . . .	212
<b>5.</b>	<b>Transformación de aleatoriedad</b>	<b>225</b>
5.1.	Paso de señales aleatorias por sistemas lineales . . . . .	225
5.1.1.	Análisis en el tiempo . . . . .	225
5.1.2.	Análisis en la frecuencia . . . . .	232
5.1.3.	Empleo de operadores lineales . . . . .	234
5.2.	Paso de señales aleatorias por sistemas no lineales . . . . .	237
5.2.1.	Métodos de linealización . . . . .	237
5.2.2.	Series de Volterra . . . . .	243
5.3.	Transformación de procesos de Markov . . . . .	249
5.3.1.	Transformación de señales aleatorias reales por procesos de Markov . . . . .	249
5.3.2.	Modelos de transformación de señales . . . . .	252
5.3.3.	Ejemplos de procesos generados de Markov . . . . .	254
5.3.4.	Discretización de procesos continuos de Markov . . . . .	257
<b>6.</b>	<b>Detección y filtración</b>	<b>265</b>
6.1.	Métodos de detección . . . . .	266
6.1.1.	Detección bayesiana de señales . . . . .	267
6.1.2.	Detección de máxima verosimilitud . . . . .	273
6.2.	Estimación de parámetros en señales aleatorias . . . . .	275
6.2.1.	Estimación en condiciones de ruido aditivo . . . . .	275
6.2.2.	Estimación de máxima verosimilitud . . . . .	275
6.2.3.	Combinación lineal de señales . . . . .	277
6.3.	Filtración óptima lineal por mínimos cuadrados . . . . .	281



---

6.3.1. Optimización de la respuesta a impulso . . . . .	281
6.3.2. Condición de realización física . . . . .	285
6.3.3. Filtros acoplados . . . . .	289
<b>7. Sistemas estocásticos . . . . .</b>	<b>295</b>
7.1. Modelos de sistemas dinámicos con entradas aleatorias . . . . .	295
7.1.1. Sistemas lineales estacionarios . . . . .	295
7.1.2. Variables de estado y modelado de procesos aleatorios . . . . .	296
7.1.3. Estimación de parámetros en sistemas estocásticos . . . . .	301
7.1.4. Identificación de parámetros en sistemas lineales . . . . .	303
7.2. Filtración de sistemas dinámicos . . . . .	307
7.2.1. Filtración Wiener . . . . .	307
7.2.2. Filtración óptima de sistemas discretos lineales . . . . .	308
7.2.3. Filtración óptima de sistemas continuos lineales . . . . .	312
7.2.4. Filtración de sistemas no lineales . . . . .	316
7.2.5. Filtración adaptativa de sistemas dinámicos . . . . .	318
<b>Bibliografía . . . . .</b>	<b>323</b>
<b>Índice alfabético . . . . .</b>	<b>327</b>

# Notaciones

Notación	Significado
$x, \mathbf{x}$	Escalar, Vector escalar
$x(s), x^*(s)$	Función con variable $s$ , Conjugado de $x$ por $s$
$x(s_k)$	Valor de la función en el argumento $s_k$
$f_{a,b}(s), \mathbf{f}$	Función $f$ con parámetros $a, b$ y argumento $s$ , vector funcional
$\{x_k : k = 1, \dots, N\}$	Serie o sucesión de valores con volumen $N$
$x[k]$	Señal discretizada (base normalizada $k$ )
$\langle x, y \rangle$	Producto interno de $x$ e $y$
$d(x_m, y_n), d(x, y)$	Distancia entre los elementos $x_m$ e $y_n$ , o entre las funciones $x$ e $y$
$\ x\ $	Norma de $x$
$x_k(t)$	Trayectoria $k$ de la señal $x(t)$
$\dot{x}(s), \ddot{x}(s)$	Primera, segunda derivada de la función $x$ por $s$
$\text{rank}(\mathbf{x}), \text{rank}(\mathbf{X})$	Rango del vector $\mathbf{x}$ , de la matriz $\mathbf{X}$
$\mathcal{E}_f, \mathcal{P}_f$	Energía de la señal $f$ , Potencia de la señal $f$
$\mathbf{Z}_{m \times n}, \mathbf{\Phi}^\top$	Matriz de orden $m \times n$ , Matriz transpuesta
$\mathbf{J}(x_1, \dots, x_n)$	Jacobiano por las variables $x_1, \dots, x_n$
$\text{supp}(x), \text{car}(x), \text{trace}(\mathbf{X})$	Soporte de la función $x$ , cardinal de $x$ , traza de $\mathbf{X}$
$\mathbf{I}, \mathbf{i}$	Matriz unitaria, vector unitario
$\mathfrak{X}$	Conjunto, espacio
$\mathcal{O}\{N\}$	Orden del número de operaciones
$o(x)$	Función de valores despreciables, $o(x)/x \rightarrow 0, x \rightarrow 0$
$\mathcal{K}\{x\}(s)$	Transformada (operador) sobre $x$ con argumento $s$
$\mathcal{F}, \mathcal{Z}, \mathcal{L}$	T. Fourier, Zeta, Laplace
$X(s)$	Representación espectral con variable $s$
$\mathbb{Z}, \mathbb{N}, \mathbb{R}, \mathbb{C}$	Dominio de los enteros, naturales, reales y complejos
$\Re\{x\}, \Im\{x\}$	Parte real de $x$ , parte imaginaria de $x$
$\xi(s)$	Señal aleatoria (alfabeto griego) de variable $s$
$\mathcal{B}(\mathfrak{X})$	$\sigma$ -álgebra $\mathcal{B}$ de subconjuntos de Borell sobre el espacio $\mathfrak{X}$
$F_\xi(x), P(x)$	Función de distribución integral, Función de probabilidad
$p_\xi(x), p_\xi(x; y)$	FDP de la variable aleatoria $\xi$ , FDP con argumentos $x$ e $y$
$\overline{E\{\xi^n\}}$	Valor esperado de orden $n$ (promedio de ensamble)
$\xi^n(t)$	Valor esperado de orden $n$ (promedio de tiempo)
$m_{n\xi}, \mu_{n\xi}$	Momento inicial, centralizado de orden $n$ de la variable $\xi$
$\sigma_\xi^2$	Varianza de la variable $\xi$
$\Theta_\xi(\omega)$	Función característica de la variable $\xi$
$\tilde{\xi}$	Valor estimado de $\xi$
$R_{\xi,\eta}(\cdot), K_{\xi,\eta}(\cdot)$	Función de correlación y covarianza mutua entre $\xi$ y $\eta$
$S_\xi(\omega)$	Espectro de potencia de $\xi$
$\Lambda_\xi$	Relación de verosimilitud para la variable aleatoria $\xi$
$\mathcal{N}, \mathcal{U}(m_{1\xi}, \sigma_\xi)$	Distribución (Gaussiana)(uniforme) con media $m_{1\xi}$ y varianza $\sigma_\xi^2$

## Acrónimos

TF	Transformada de Fourier
TIF	Transformada Inversa de Fourier
DEE	Densidad espectral de energía
DEP	Densidad espectral de potencia
TFTC	Transformada de Fourier en tiempo corto
TL	Transformada de Laplace
TW	Transformada wavelet
FDP	Función densidad de probabilidad
K-L	Karhunen-Loève (Transformada)
RBG	Ruido blanco Gaussiano



# Capítulo 1

## Representación de señales y sistemas

La representación, entendida como una idea que sustituye a la realidad, en el caso de señales y sistemas se orienta a la descripción de sus principales propiedades de interacción. En este capítulo se analizan los métodos de representación discreta e integral como formas fundamentales de la descripción de señales y sistemas.

### 1.1. Representación discreta de señales

**Definición 1.1.** Un espacio métrico corresponde a la pareja  $\mathfrak{X} = (X, d)$ , que consta del conjunto  $X$  o espacio de elementos (puntos) y la distancia  $d$ , que es una función real positiva semidefinida, que para cualquier pareja de elementos  $x, y \in X$ , cumple los axiomas:

1.  $d(x, y) = 0$ , si y solo si,  $x = y$ ; elementos con propiedades geométricas iguales.
2.  $d(x, y) = d(y, x)$ ; axioma de simetría
3.  $d(x, z) \leq d(x, y) + d(y, z)$ ; axioma triangular

#### 1.1.1. Espacio de representación de señales

En general, un conjunto de señales de análisis es el constituido por todas las funciones de valor complejo definidas en forma continua sobre un eje real, por ejemplo el del tiempo,  $L = \{x = x(t) : x(t) \in \mathbb{C}, t \in \mathbb{R}\}$ , donde  $L = \{x : \mathcal{K}\}$  es el conjunto formado por todos los  $x$ , para los cuales el atributo o relación  $\mathcal{K}$  es cierto:  $\mathcal{K} \Rightarrow x \in L$ . La mayoría de los espacios de funciones de señales se restringen a los espacios convencionales de Lebesgue:

$$L^p = L^p(\mathbb{R}) = \left\{ x \in L : \|x\|_p = \left( \int_{\mathbb{R}} |x(t)|^p dt \right)^{1/p} < \infty \right\}, p \geq 1 \quad (1.1)$$
$$L^\infty = L^\infty(\mathbb{R}) = \left\{ x \in L : \|x\|_\infty = \sup_t |x(t)| < \infty \right\}$$

donde  $\|\mathbf{x}\|_p$  es la *norma* definida para  $\mathbf{x}$  en el espacio  $L^p$ . Un espacio lineal  $L$ , en el cual se define una norma, se denomina *normalizado*. La restricción de  $p \geq 1$ ,  $p \in \mathbb{Z}$ , en (1.1), implica que la clase  $L^p(\mathbb{R})$  es un espacio normalizado, el cual es *completo* con respecto a la respectiva norma, esto es, cualquier sucesión fundamental de elementos converge.

La generación de espacios de señales, a partir de alguna condición común con interpretación física (energía, transformación a algún espacio complejo, etc.), implica establecer el modelo matemático formal de relación  $\mathcal{X}$  entre los elementos del conjunto. Un conjunto de funciones provisto de una distancia adecuada, conforma un *espacio de señales*. Un ejemplo de distancia entre dos señales,  $\mathbf{x}(t)$ ,  $\mathbf{y}(t) \in \mathbb{C}$ , definidas en  $T$ , es:

$$d(\mathbf{x}, \mathbf{y}) = \|\mathbf{x} - \mathbf{y}\|_2 = \left( \int_T |\mathbf{x}(t) - \mathbf{y}(t)|^2 dt \right)^{1/2} \quad (1.2)$$

El conjunto de funciones relacionados por la distancia (1.2), denominado espacio con *métrica cuadrática*, para los cuales la respectiva norma sea acotada,  $\|\mathbf{x}\|_2 < \infty$ , o *espacio*  $L^2(T)$  que es un espacio de Hilbert, está provisto del siguiente producto interno:

$$\langle \mathbf{x}, \mathbf{y} \rangle_{L^2(T)} = \int_T \mathbf{x}^*(t) \mathbf{y}(t) dt$$

El espacio de funciones, dado por la distancia (1.2), tiene amplio uso en la representación de señales debido a la interpretación física simple de su norma, la cual se asocia a la energía de las señales. Cuando  $\mathbf{x} \in L^2(T)$ , se tiene que,  $\|\mathbf{x}\|_2 = \langle \mathbf{x}, \mathbf{x}^* \rangle_{L^2(T)} \triangleq \mathcal{E}_x$ , que es la *energía* de la señal. Por cierto, la señal  $\mathbf{x}(t)$  se denomina *señal de energía* cuando

$$\mathcal{E}_x = \|\mathbf{x}\|_2 = \int_T |\mathbf{x}(t)|^2 dt < \infty, \quad (1.3)$$

La representación discreta de una señal  $\mathbf{x} \in L^2(T)$  de energía, que cumpla (1.3), implica hallar la transformación del espacio  $L^2(T)$  en el espacio  $\mathbb{C}^n$ , donde el valor de  $n$  se elige a partir del compromiso entre la precisión y la economía de la representación. La forma general para hallar esta representación consiste en la selección de un subespacio con dimensión  $n$  a partir de  $L^2(T)$ . Teniendo en cuenta que  $L^2(T)$  es un espacio completo separable [1], la señal  $\mathbf{x} \in L^2(T)$  puede ser representada de manera aproximada con cualquier precisión, si la dimensión de representación se escoge suficientemente grande, por medio de un conjunto de valores o coeficientes  $\mathbf{x}_k$ , ( $k \rightarrow \infty$ ), expresados en combinación lineal del siguiente espacio de funciones ordenadas, que se elige adecuadamente:

$$\mathbf{x}(t) = \sum_k^{\infty} \mathbf{x}_k \phi_k(t), \quad (1.4)$$

donde  $\phi_k(t)$  es el conjunto de funciones elegidas a priori, que conforman una base del espacio vectorial  $L^2(T)$ , las cuales son denominadas *funciones base*.

### 1.1.2. Descomposición en funciones ortogonales

**Definición 1.2.** Un espacio de señales complejas  $\{\phi_k(t) \in \mathbb{C}\}$  se define *ortogonal*, dentro del intervalo de representación  $(t_i, t_f)$ , si para el producto interno  $\langle \phi_m, \phi_n^* \rangle$  se cumple:

$$\langle \phi_m, \phi_n^* \rangle = \int_{t_i}^{t_f} \phi_m(t) \phi_n^*(t) dt = \int_{t_i}^{t_f} \phi_m^*(t) \phi_n(t) dt = \begin{cases} 0, & m \neq n \\ \mathcal{E}_m, & m = n \end{cases} \quad (1.5)$$

siendo  $\mathcal{E}_m = \mathcal{E}_n = \|\phi\|_2$  un valor de energía.

Como se deduce de la expansión ortogonal (1.4), una señal de energía  $x(t)$  se puede representar de forma aproximada en términos de un conjunto de funciones base  $\{\phi_n(t) : n = 0 \dots, N\}$ , para el cual los coeficientes  $x_n$  caracterizan el peso de la correspondiente función ortogonal  $\phi_n(t)$ . Los valores  $x_n$  se determinan de acuerdo con la condición del mínimo error cuadrático medio:

$$\overline{\varepsilon_N^2(t)} = \frac{1}{t_f - t_i} \int_{t_i}^{t_f} \left| x(t) - \sum_{n=0}^N x_n \phi_n(t) \right|^2 dt, \text{ siendo } \overline{\varepsilon_N^2(t)} \geq 0 \quad (1.6)$$

Como resultado se obtienen los siguientes coeficientes de descomposición:

$$x_n = \frac{\langle x, \phi_n \rangle}{\|\phi\|_2} = \frac{\int_{t_i}^{t_f} x(t) \phi_n^*(t) dt}{\int_{t_i}^{t_f} |\phi_n(t)|^2 dt} = \frac{1}{t_f - t_i} \int_{t_i}^{t_f} x(t) \phi_n^*(t) dt \quad (1.7)$$

Una señal de potencia (o energía) se descompone mediante un conjunto de funciones base, que genera un espacio completo, cuando los coeficientes  $x_n$  se determinan por (1.7).

El conjunto de los coeficientes  $\{x_n\}$  se denomina *espectro de la señal  $x(t)$* , mientras cada producto  $x_n \phi_n(t)$  se define como su respectiva componente espectral. Cualquiera de las dos formas, la serie generalizada de Fourier (1.4) o *el espectro  $\{x_n\}$*  (1.7), determinan unívocamente la señal  $x(t)$ .

### 1.1.3. Ejemplos de conjuntos ortogonales completos

El conjunto ortogonal más conocido corresponde a la *serie exponencial de Fourier*, que se define por las funciones exponenciales complejas del tipo:

$$\phi_n(t) = e^{jn\omega_0 t}, \quad (1.8)$$

donde  $\omega_0 = 2\pi/\mathcal{E}_n$ , siendo  $\mathcal{E}_n = (t_f - t_i)$  y  $n \in \{0, \pm 1, \pm 2, \dots\}$ , que corresponde a cada término de la serie o *armónico*.

Cabe anotar, que todas las expresiones anteriores también son válidas para la representación espectral de las funciones periódicas, reemplazando el intervalo de definición de las señales aperiódicas  $[t_i, t_f]$  por el valor del período  $T$  de las funciones periódicas.

En la Tabla (1.1) se muestran los coeficientes de Fourier para algunas señales comunes.

Señal	Definición	Coficiente $x_n$
Cuadrada simétrica	$\begin{cases} 1, &  t  < \frac{T}{4} \\ -1, & \frac{T}{4} <  t  < \frac{T}{2} \end{cases}$	$\begin{cases} \text{sinc}(n\pi/2), & n \neq 0 \\ 0, & n = 0 \end{cases}$
Pulsos rectangulares	$\begin{cases} +1, &  t  < \frac{\tau}{2} \\ 0, & \frac{\tau}{2} \leq  t  < \frac{T}{2} \end{cases}$	$\frac{\tau}{T} \text{sinc}(n\pi\tau/T)$
Triangular simétrica	$1 - 4 t /T, \quad  t  < T/2$	$\begin{cases} \text{sinc}(n\pi/2), & n \neq 0 \\ 0, & n = 0 \end{cases}$

**Tabla 1.1.** Coeficientes de descomposición espectral de Fourier

Se considera que los motivos principales, por los cuales las series de Fourier tienen amplia aplicación, son los siguientes [2]:

- Se representan mediante oscilaciones simples determinadas para todo valor de  $t$ .
- Se descomponen en oscilaciones armónicas **sin** y **cos**, que son las únicas funciones del tiempo que conservan su forma al pasar por cualquier dispositivo lineal; sólo varían su fase y su amplitud.
- La descomposición en senos y cosenos permite aplicar el método simbólico usualmente desarrollado para el análisis de circuitos lineales.

En la aproximación de señales continuas se emplean diferentes polinomios y funciones especiales ortogonales. Escogiendo adecuadamente estas funciones en la descomposición ortogonal (1.4), se obtiene la representación aproximada para la señal a partir de un número pequeño de elementos de la serie.

De otra parte existen funciones, que conforman sistemas base completos  $\{\phi_n(t)\}$ , siendo ortonormales en los intervalos indicados con valor de peso  $\sqrt{w(t)}$ , esto es, que se cumpla la condición:

$$\{\phi_n(t)\} = \left\{ \sqrt{w(t)}\psi_n(t) \right\}, \quad \forall n \quad (1.9)$$

donde  $w(t)$  es la *función de peso* y  $\{\psi_n(t)\}$  son las funciones sobre los cuales se forman los sistemas base ortogonales. Como consecuencia de (1.9), la definición de ortogonalidad (1.5) sobre el intervalo  $T$ , cambia por:

$$\int_T \psi_m(t) \psi_n^*(t) w(t) dt = \begin{cases} 0, & m \neq n \\ \mathcal{E}_n, & m = n \end{cases}, \quad \mathcal{E}_n = \left\| \sqrt{w(t)}\psi_n(t) \right\|_2 = \int_T \left( \sqrt{w(t)}\psi_n(t) \right)^2 dt$$



Así mismo, cambia la definición de los coeficientes (1.7) de la expansión (1.4):

$$x_n = \frac{1}{\left\| \sqrt{w(t)} \psi_n(t) \right\|_2} \int_T x(t) \sqrt{w(t)} \psi_n^*(t) dt. \quad (1.10)$$

Particular importancia tienen las funciones de Walsh, para las cuales  $\kappa(l)$  corresponde al elemento  $l$  en representación del valor  $\kappa$  en código Gray, mientras  $r_n(t)$  son las funciones de Rademascher [3]. La descomposición de una señal,  $x(t)$ , en este caso tiene la forma:

$$x(t) = \sum_{n=0}^{\infty} x_n \text{wal}_n(t) \quad (1.11)$$

donde los coeficientes de la serie Fourier-Walsh (1.11), que se determinan de acuerdo con (1.7) y (1.10),

$$x_n = \int_{-1/2}^{1/2} x(t) \text{wal}_n(t) dt$$

conforman el espectro de la señal  $x(t)$  por la base Walsh, el cual es denominado *S-espectro* (*sequency spectrum*).

**Ejemplo 1.1.** Determinar el *S*-espectro de la señal  $x(t) = \sin 2\pi t$ ,  $-\infty < t < \infty$ , empleando el conjunto  $\{\text{wal}_n(t), 0 \leq t < 1, n = 0, 1, \dots, 31\}$ .

Debido a la imparidad de la función  $x(t)$  respecto al punto  $t = 1/2$ , todos los coeficientes para las funciones pares  $\text{wal}_{2n}(t) = 0$ , al igual que los coeficientes de las funciones impares 3, 7, 11, 15, 19, 23, 27, 31. Los demás coeficientes tienen los siguientes valores:

$$\begin{aligned} x_1 &= \int_{-1/2}^{1/2} x(t) \text{wal}_1(t) dt = \int_{-1/2}^{1/2} \sin(2\pi t) \text{wal}_1(t) dt = 2 \int_0^{1/2} \sin(2\pi t) dt = \frac{2}{\pi} \approx 0.637 \\ x_5 &= \int_{-1/2}^{1/2} \sin(2\pi t) \text{wal}_5(t) dt = 4 \int_0^{1/8} \sin(2\pi t) dt - 2 \int_{1/8}^{3/8} \sin(2\pi t) dt \approx -0.264 \end{aligned}$$

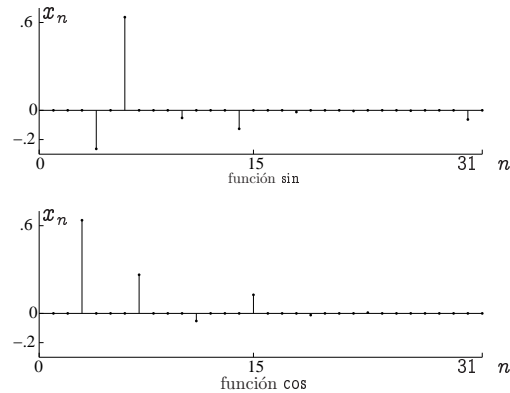
En forma similar, se obtienen los siguientes coeficientes:

$$\begin{aligned} x_9 &\approx -0.052, \quad x_{13} \approx -0.127, \quad x_{17} \approx -0.012, \quad x_{21} \approx -0.005, \\ x_{25} &\approx -0.026, \quad x_{29} \approx -0.063 \end{aligned}$$

De igual manera, se puede establecer el *S*-espectro de la señal  $x(t) = \cos 2\pi t$ , para la cual se obtiene los siguientes coeficientes:  $x_2 \approx 0.637$ ,  $x_6 \approx 0.264$ ,  $x_{10} \approx -0.052$ ,  $x_{14} \approx 0.127$ ,  $x_{18} \approx -0.012$ ,  $x_{22} \approx 0.005$ . Los demás coeficientes son iguales a cero. En la Figura 1.1 se observan las respectivas representaciones.

En general, el *S*-espectro de la función  $x(t) = \sin(2\pi t + \theta)$  se obtiene a partir de la relación:

$$\sin(2\pi t + \theta) = \sin(2\pi t) \cos \theta + \cos(2\pi t) \sin \theta$$



**Figura 1.1.** Representación espectral por la base Walsh

por lo cual es claro, que cambiando el desfase  $\theta$  cambian los valores de los coeficientes de descomposición de la serie de Walsh.

## Problemas

**Problema 1.1.** Sean las señales  $y(t)$  y  $x(t)$  elementos de un espacio real de Hilbert, tales que son linealmente independientes, y por tanto la igualdad  $y = \lambda x$  no tiene lugar para ningún valor real de  $\lambda$ . Demostrar la desigualdad de Cauchy-Buniakovski,  $|\langle x, y \rangle| < \|x\| \cdot \|y\|$ .

**Problema 1.2.** En un espacio de Hilbert están dados los vectores  $u$  y  $v$ , tal que  $\|v\| = 1$ . En analogía con la geometría de los vectores comunes en un plano, el vector  $w = \langle u, v \rangle$  se denomina *proyección ortogonal* del vector  $u$  en la dirección  $v$ . Demostrar que el vector  $y = u - w$  es ortogonal al vector  $v$ .

**Problema 1.3.** Demostrar la ortogonalidad del conjunto formado por el principio  $\text{sgn}[\sin(2^m \pi \theta)]$ ,  $m \in \mathbb{Z}$ .

**Problema 1.4.** Una señal compleja periódica  $x(t)$  en el intervalo de tiempo  $-T/2 \leq t \leq T/2$  tiene la forma,  $x(t) = x_1(t) + jx_2(t)$ . Demostrar que si la función  $x_1(t)$  es par, mientras la función  $x_2(t)$  es impar, entonces los coeficientes  $x_n$  de la serie de Fourier, para cualquier  $n$ , serán valores reales.

**Problema 1.5.** Calcular los coeficientes de la serie compleja de Fourier para las siguientes funciones periódicas dadas en el intervalo  $[-\pi, \pi]$ :

a).  $x(t) = t$ . b).  $x(t) = |\sin t|$ . c).  $x(t) = \text{sgn}(t)$ .

**Problema 1.6.** Determinar el espectro de la serie Haar-Fourier para la sucesión periódica de pulsos cuadrados con amplitud  $a = 2^m$ ,  $m \in \mathbb{Z}$ , apertura  $\tau = 1/a$ , y periodo unitario, empleando el sistema  $\{\text{har}_n(t), 0 \leq t < 1, n = 0, 1, 2, \dots, N, N = 2^k - 1, k \in \mathbb{Z}\}$  siendo  $m = 3$  y  $k = 4$ .

**Ejercicio en el CP 1.1.** Desarrollar un programa gráfico que convierta la serie trigonométrica de Fourier utilizando desde 1 hasta 15 armónicos y 50 puntos por periodo.

## 1.2. Representación integral de señales y sistemas

### 1.2.1. Densidad espectral

La representación discreta, analizada en el numeral §1.1, sirve en el caso de descomposición de señales periódicas o de las señales aperiódicas, definidas sobre intervalos finitos de tiempo. Esta última restricción es superable al emplear la representación integral, la cual resulta en forma de transformadas integrales, que se lleva a cabo para conjuntos base continuos. Específicamente, la generalización de (1.4), para cualquier conjunto base o *transformada directa generalizada de Fourier*, se puede escribir en el dominio  $T$  como sigue:

$$v(s) = \int_T x(t) \theta(s, t) dt, \quad s \in S \quad (1.12)$$

siendo  $s$  el parámetro generalizado del dominio  $S$  y  $\theta(s, t)$  el *núcleo base conjugado*. La función  $v(s)$ , que es una representación continua de  $x(t)$  similar a los coeficientes  $x_n$  en (1.4), corresponde a la *densidad*, la cual caracteriza la distribución de  $x(t)$  con relación a  $\theta(s, t)$  en las diferentes regiones de  $S$ . De forma similar, se determina la *transformada inversa generalizada de Fourier*:

$$x(t) = \int_S v(s) \phi(s, t) ds, \quad t \in T \quad (1.13)$$

Al reemplazar  $x(t)$  de (1.13) en (1.12) e intercambiando el orden de integración se obtiene la condición conjunta que cumplen los núcleos base conjugados:

$$x(t) = \int_S \int_T x(\tau) \theta(s, \tau) \phi(t, s) d\tau ds = \int_T k(t, \tau) x(\tau) d\tau = f_k \{x(t)\} \quad (1.14)$$

donde el núcleo de la integral se define como:

$$k(t, \tau) = \int_S \theta(s, \tau) \phi(t, s) ds$$

En forma general, el funcional lineal  $f_k \{x(t)\}$  no es acotado ni continuo, por lo tanto no es de esperar que exista una función, perteneciente a  $L^2(T)$  y definida en  $\tau$ , que cumpla la condición (1.14) para cualquier señal  $x(t)$ . La solución a este problema está en la representación de  $x(t)$  mediante la función generalizada singular  $\delta(t)$  de la forma [4]:

$$f_k \{x(t)\} = x(t) = \int_T \delta(t - \tau) x(\tau) d\tau, \quad t \in T$$

Dado un núcleo base, de (1.14) se establece la restricción para su núcleo conjugado:

$$\int_S \phi(t, s) \theta(s, \tau) ds = \delta(t - \tau) \quad (1.15)$$

De la misma manera, al reemplazar (1.12) en (1.13) se tiene la condición:

$$\int_T \theta(s, t) \phi(t, \sigma) dt = \delta(s - \sigma) \quad (1.16)$$

que debe cumplir tanto  $\theta(s, t)$ , como  $\phi(t, \sigma)$  para conformar un par de núcleos conjugados base. Sin embargo, el análisis de (1.15) y (1.16) muestra que ambos núcleos al mismo tiempo no pueden pertenecer a  $L^2(\mathbb{R})$ . Algunos núcleos base tienen la propiedad por la cual la transformada de la función en el dominio  $t \in T$ , siendo integrable y perteneciente a  $L^2(T)$ , siempre genera funciones en el dominio  $s \in S$  también integrables (pertenecientes a  $L^2(S)$ ). La correspondencia completa entre las funciones determinadas en los dominios de  $t$  y  $s$  se provee mediante núcleos autoconjugados (también denominados *inversos o mutuos*), esto es [4],  $\theta^*(s, t) = \phi(t, s)$ .

La transformada integral de Fourier, o *Transformada de Fourier* (TF), se obtiene asumiendo en calidad de base (1.8) la función,  $\phi(t, s) = e^{j2\pi st}$  a la cual le corresponde el núcleo base conjugado,  $\theta(t, s) = e^{-j2\pi st}$ , definido para los argumentos  $T \subset (-\infty, \infty)$  y  $S \subset (-\infty, \infty)$ , que corresponde a la *Transformada Inversa de Fourier* (TIF):

$$x(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} X(\omega) e^{j\omega t} d\omega \triangleq \mathcal{F}^{-1}\{X(\omega)\} \quad (1.17)$$

De igual manera, se puede obtener la *Transformada Directa de Fourier* (TDF):

$$X(\omega) = \int_{-\infty}^{\infty} x(t) e^{-j\omega t} dt \triangleq \mathcal{F}\{x(t)\} \quad (1.18)$$

El par de transformadas (1.17) y (1.18) representa la señal  $x(t)$  como una suma continua de funciones exponenciales cuyas frecuencias están en el intervalo  $(-\infty, \infty)$ .

La transformada de Fourier  $X(\omega)$  corresponde a la función de *densidad espectral de energía* (DEE), cuya existencia depende de las *condiciones de Dirichlet* [5]:

1.  $x(t)$  tiene un número finito de máximos y mínimos en cualquier intervalo de tiempo finito,
2.  $x(t)$  tiene un número finito de discontinuidades finitas en cualquier intervalo de tiempo finito,
3.  $x(t)$  es absolutamente integrable.

Con relación a la última condición, se puede decir que la transformada de Fourier representa unívocamente cualquier señal real de energía. Así mismo, se considera que las condiciones de Dirichlet son suficientes para la convergencia de la TF en la representación de una señal de potencia en el intervalo  $(0, T)$ .

Las principales propiedades de la TF se muestran en la Tabla 1.2.

Propiedad	Descripción
Linealidad	$\mathcal{F}\left\{\sum_n a_n x_n(t)\right\} = \sum_n a_n X_n(\omega), \forall a_n = \text{const.}$
Conjugada compleja	$\mathcal{F}\{x^*(t)\} = X^*(-\omega)$
Dualidad	$\mathcal{F}\{X(t)\} = 2\pi x(-\omega)$
Escala de coordenadas	$\mathcal{F}\{x(\alpha t)\} = \frac{X(\omega/\alpha)}{ \alpha }$
Desplazamiento en el tiempo (retardo)	$\mathcal{F}\{x(t - t_0)\} = X(\omega) e^{-j\omega t_0}$
Desplazamiento de frecuencia (modulación)	$\mathcal{F}\{x(t) e^{\pm j\omega_0 t}\} = X(\omega \mp \omega_0)$
Diferenciación	$\mathcal{F}\left\{\frac{d}{dt}x(t)\right\} = j\omega X(\omega)$
Integración	$\mathcal{F}\left\{\int_{-\infty}^t x(\tau) d\tau\right\} = \frac{1}{j\omega}X(\omega) + \pi X(0)\delta(\omega)$
Teorema de Rayleigh	$\mathcal{E}_x = \int_{-\infty}^{\infty} X(\omega) X(-\omega) d\omega = \frac{1}{2\pi} \int_{-\infty}^{\infty}  X(\omega) ^2 d\omega$
Teorema de Parsevall	$\mathcal{E}_x = \int_{-\infty}^{\infty} X(\omega) X^*(\omega) d\omega$ $= \frac{1}{2\pi} \int_{-\infty}^{\infty}  X(\omega) ^2 d\omega = 2\pi \int_{-\infty}^{\infty} x^2(t) dt$
Multiplicación	$\mathcal{F}\{x_1(t) x_2(t)\} = X_1(\omega) * X_2(\omega)$
Convolución	$\mathcal{F}\{x_1(t) * x_2(t)\} = X_1(\omega) X_2(\omega)$

Tabla 1.2. Propiedades de la transformada de Fourier

**Ejemplo 1.2.** Determinar la DEE de la función pulso cuadrado  $x(t) = a \text{rect}_\tau(t)$ .

Al realizar el cálculo de la TDF, (1.18), se obtiene:

$$\begin{aligned}
 X(\omega) &= \int_{-\infty}^{\infty} a \text{rect}_\tau(t) e^{-j\omega t} dt = a \int_{-\tau/2}^{\tau/2} e^{-j\omega t} dt = \frac{a}{j\omega} (e^{j\omega\tau/2} - e^{-j\omega\tau/2}) = \frac{a\tau \sin(\omega\tau/2)}{\omega\tau/2} \\
 &= a\tau \text{sinc}(\omega\tau/2).
 \end{aligned}$$

En general, se pueden formar nuevos conjuntos base a partir del desplazamiento en el tiempo de un sistema base original, así

$$\phi(s, t) = \phi(t - s) \quad (1.19)$$

entonces, el sistema base es función de una sola variable, dada por la diferencia de ambos argumentos,  $(t - s)$ .

Si se considera  $T, S \in (-\infty, \infty)$ , entonces, la señal original es la *integral de convolución*:

$$x(t) = \int_{-\infty}^{\infty} v(s) \phi(s-t) ds \triangleq v(t) * \phi(t) \quad (1.20)$$

El núcleo conjugado, en caso de existir, se puede hallar mediante la TF de (1.20), que aplicando la propiedad de convolución en el tiempo, se obtiene que  $X(\omega) = V(\omega) \Phi(\omega)$ , de lo que resulta que  $V(\omega) = \Theta(\omega) X(\omega)$ , donde  $\Theta(\omega) = 1/\Phi(\omega)$ . Por lo tanto, el núcleo conjugado también depende de la diferencia de argumentos

$$v_x(s) = \int_{-\infty}^{\infty} x(t) \theta(t-s) dt \quad (1.21)$$

En calidad de ejemplo concreto, en (1.19), se puede analizar el desplazamiento de la función base  $\phi(t-s) = \delta(t-s)$ , entonces  $\Phi(f) = \Theta(f) = 1$ , con lo cual se obtiene que

$$v_x(s) = \int_{-\infty}^{\infty} x(t) \delta(t-s) dt = x(s) \quad (1.22)$$

**Transformada de Laplace.** El análisis de algunas señales  $x(t)$ , mediante la TF se dificulta cuando no cumplen la condición débil de Dirichlet, esto es, no son absolutamente integrables. En este caso, se puede hallar la convergencia, si la señal se multiplica por una función que acote los valores extremos, por ejemplo, por una curva exponencial del tipo,  $e^{-ct}$ , en la cual la constante  $c > 0$  se escoge de tal manera que asegure la condición de integrabilidad absoluta del producto  $x(t)e^{-ct}$ , entonces, la DEE, (1.17), toma la forma,

$$X(\omega, c) = \int_{-\infty}^{\infty} (x(t) e^{-ct}) e^{-j\omega t} dt \quad (1.23)$$

Sin embargo, para asegurar la convergencia de (1.23), se debe ajustar la señal en el tiempo,  $x(t)$ , tal que  $\forall t < 0$ , su aporte sea 0. De otra manera, el multiplicando  $e^{-ct}$  puede conllevar a la divergencia de la integral. Por lo tanto, el límite inferior de la densidad acotada siempre es 0:

$$X(\omega, c) = \int_0^{\infty} (x(t) e^{-ct}) e^{-j\omega t} dt = \int_0^{\infty} x(t) e^{-(c+j\omega)t} dt = X(c+j\omega) \quad (1.24)$$

Si se halla la TF inversa de la densidad espectral  $X(c+j\omega)$ , se obtiene que

$$\frac{1}{2\pi} \int_{-\infty}^{\infty} X(c+j\omega) e^{j\omega t} d\omega = x(t) e^{-ct}$$

Así mismo, si ambas partes de la integral anterior se multiplican por  $e^{ct}$ , juntando los factores de multiplicación exponencial y haciendo el cambio de variable  $\omega \rightarrow c + j\omega$ , entonces, se llega a la expresión:

$$\frac{1}{2\pi j} \int_{c-j\infty}^{c+j\infty} X(c + j\omega) e^{(c+j\omega)t} d(c + j\omega) = x(t)$$

En la práctica, es usual el empleo de la notación  $p \triangleq c + j\omega$ , entonces la anterior integral y la densidad (1.24), conforman el par de *Transformadas de Laplace* (TL):

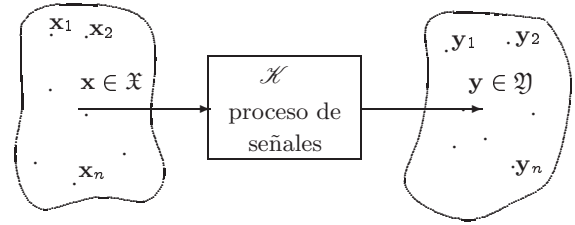
$$\int_0^{\infty} x(t) e^{-pt} dt = X(p) \triangleq \mathcal{L}\{x(t)\} \quad (1.25a)$$

$$\frac{1}{2\pi j} \int_{c-j\infty}^{c+j\infty} X(p) e^{pt} dp = x(t) \triangleq \mathcal{L}^{-1}\{X(p)\} \quad (1.25b)$$

De (1.25b), se observa que la TL inversa se realiza mediante la integración en el plano complejo  $p$  sobre toda la recta  $c = \text{const.}$  También es claro, que el reemplazo de  $p = j\omega$  conduce directamente a la TF, por lo que, todas sus propiedades tienen lugar para la TL.

### 1.2.2. Transformaciones lineales

A su paso por los sistemas, las señales sufren cambios en cada una de las etapas del proceso. Debido a la diversidad de formas posibles de señal es preferible tener una representación común para su análisis. La transformación de la señal se puede analizar como el enlace entre la entrada y la salida de un sistema o circuito (Figura 1.2), pues en sí corresponde a la representación de un conjunto inicial de señales  $\mathfrak{X}$  en términos de otro conjunto; el de salida  $\mathfrak{Y}$ .



**Figura 1.2.** Transformación de señales

En forma general, para la descripción de la transformación  $\mathcal{K}$  se deben conocer todas las posibles parejas ordenadas entrada - salida,  $\{\mathbf{x}, \mathcal{K}\{\mathbf{x}\}\}$ , lo cual en la mayoría de los casos prácticos no es realizable, entre otras razones, porque se puede tener una alta cantidad de enlaces, además, por la dificultad en su ordenamiento. Sin embargo, cuando se restringe el análisis a las transformadas lineales, la transformación se puede describir completamente empleando un subconjunto reducido de todas las posibles parejas entrada - salida, como consecuencia de la propiedad de superposición de las transformadas lineales. Los modelos lineales son el referente de sistemas más empleados, de hecho, es común el análisis de sistemas no lineales, a través de su linealización.

Una transformación lineal  $\mathcal{K}$ , que corresponde a la imagen determinada en un espacio lineal  $\mathfrak{X}$ , tiene las siguientes propiedades:

1. *Aditividad*,  $\mathcal{K}\{\mathbf{x}_1 + \mathbf{x}_2\} = \mathcal{K}\{\mathbf{x}_1\} + \mathcal{K}\{\mathbf{x}_2\}$ ,  $\forall \mathbf{x}_1, \mathbf{x}_2 \in \mathfrak{X}$ ,
2. *Homogeneidad*,  $\mathcal{K}\{\alpha \mathbf{x}\} = \alpha \mathcal{K}\{\mathbf{x}\}$ ,

o lo que es equivalente  $\mathcal{K}\{\alpha \mathbf{x}_1 + \beta \mathbf{x}_2\} = \alpha \mathcal{K}\{\mathbf{x}_1\} + \beta \mathcal{K}\{\mathbf{x}_2\}$ , donde  $\alpha, \beta = \text{const.}$  De lo anterior resulta que  $\mathcal{K}\{\mathbf{0}\} = \mathbf{0}$  y  $\mathcal{K}\{-\mathbf{x}\} = -\mathcal{K}\{\mathbf{x}\}$ , por lo tanto, el conjunto de vectores linealmente transformados corresponde a un espacio lineal con el mismo conjunto de escalas que el conjunto de definición de la transformada.

De otra parte, si los espacios de entrada y salida son idénticos, la transformación lineal se denomina *operador lineal*, esto es, cuando el dominio de representación original coincide con el dominio de la imagen de la transformada. En los operadores lineales la multiplicación de dos operadores se determina como:

$$\mathcal{K} = \mathcal{K}_1 \mathcal{K}_2 \Rightarrow \mathcal{K}\{\mathbf{x}\} = \mathcal{K}_1\{\mathcal{K}_2\{\mathbf{x}\}\}$$

que corresponde a la conexión en cascada de ambos operadores. Cabe anotar que el conjunto  $\mathbf{x} \in \mathfrak{X}$ , tal que  $\mathcal{K}\{\mathbf{x}\} = \mathbf{0}$ , se denomina núcleo o *kernel* ( $\ker \mathcal{K}$ ) del operador  $\mathcal{K}$ .

### Transformaciones lineales sobre espacios finitos

Un caso de análisis frecuente corresponde a las transformaciones lineales sobre subespacios funcionales de  $L^2(T)$  cuyos resultados se pueden generalizar a todo  $L^2(\mathbb{R})$ .

**Representación con vectores respuesta.** Sea el espacio de entrada  $\mathfrak{X}$  generado por la base linealmente independiente  $\{\varphi_i : i = 1, \dots, n\}$ , para la cual existe la base conjugada con elementos  $\{\theta_i : i = 1, 2, \dots, n\}$ . Sea el espacio imagen  $\mathfrak{Y}$  con dimensión  $n$  que contiene los valores de la transformación  $\mathcal{K}$ . Luego, a partir de la expansión generalizada de Fourier,  $\forall \mathbf{x} \in \mathfrak{X}$ , se obtiene que  $\mathbf{x}(t) = \sum_{i=1}^n \langle \mathbf{x}, \theta_i \rangle \varphi_i(t)$ ,  $t \in T$ , por lo que al considerar la linealidad de  $\mathcal{K}$ , entonces

$$\mathbf{y}(t) = \mathcal{K}\{\mathbf{x}(t)\} = \sum_{i=1}^n \langle \mathbf{x}, \theta_i \rangle \psi_i(t) \quad (1.26)$$

donde  $\{\psi_i : \psi_i(t) = \mathcal{K}\{\varphi_i(t)\}; i = 1, \dots, n\}$ , que son las respuestas para cada una de las funciones base del espacio  $\mathfrak{X}$ , dadas en calidad de entrada de  $\mathcal{K}$ . Luego, se puede considerar el conjunto  $\{\psi_i\}$  como la representación de  $\mathcal{K}$  en términos de la base  $\{\phi_i\}$  en  $\mathfrak{X}$ .

**Representación por sucesión de funcionales lineales.** Otra forma de representación de  $\mathcal{K}$  se da por la sucesión ordenada de  $n$  vectores en  $\mathfrak{X}$ . Sea  $\{\widehat{\varphi}_i : i = 1, \dots, n\}$  el conjunto de vectores linealmente independientes, que generan a  $\mathfrak{Y}$ , a los cuales le corresponde la base



conjugada  $\{\widehat{\theta}_i : i = 1, \dots, n\}$ . Entonces, cualquier vector de  $\mathfrak{Y}$  se representa como:

$$y(t) = \sum_{j=1}^n \langle \mathbf{y}, \widehat{\theta}_j \rangle \widehat{\varphi}_j(t) \quad (1.27)$$

teniendo en cuenta (1.26) y (1.27), se obtiene

$$y(t) = \mathcal{K} \{x(t)\} = \sum_{j=1}^n \langle \mathbf{x}, \phi_j \rangle \widehat{\varphi}_j(t), \quad \phi_i(t) = \sum_{j=1}^n \langle \widehat{\theta}_j, \psi_i \rangle \theta_i(t)$$

Por lo tanto, la sucesión ordenada  $\{\phi_j\}$  representa a  $\mathcal{K}$  en términos de la base dada de  $\mathfrak{Y}$ . Asumiendo la independencia lineal de  $\{\phi_j\}$ , entonces, la representación del operador se puede entender como una sucesión ordenada, en la cual cada elemento es una transformada de rango unitario (funcional lineal)  $f_j(\widehat{\mathbf{x}}) = \langle \mathbf{x}, \phi_j \rangle$ , con  $j = 1, \dots, n$ ; la suma por todos elementales representa toda la transformación:  $\mathcal{K} \{x(t)\} = \sum_{j=1}^n f_j(\mathbf{x}) \widehat{\varphi}_j(t)$ .

**Representación matricial.** Por cuanto los vectores  $\psi_i$  en el primer método, ó  $\phi_j$  en el segundo método, se describen por un conjunto de  $n$  coeficientes (vector fila) la transformación lineal puede ser representada por una tabla de relación de  $n \times n$  escalares, esto es, en forma de un arreglo matricial.

Sea

$$\psi_i = \sum_{j=1}^n \lambda_{ji} \widehat{\varphi}_j(t), \quad \lambda_{ji} = \langle \psi_i, \widehat{\theta}_j \rangle = \langle \mathcal{K} \{\varphi_i\}, \widehat{\theta}_j \rangle \quad (1.28)$$

Luego, al reemplazar (1.28) en (1.26) se tiene que:

$$y(t) = \sum_{i=1}^n \sum_{j=1}^n \lambda_{ji} \alpha_i \varphi_j(t) = \sum_{j=1}^n \beta_j \widehat{\varphi}_j(t), \quad (1.29)$$

donde  $\alpha_i = \langle \mathbf{x}, \theta_i \rangle$ , y  $\beta_i = \langle \mathbf{y}, \widehat{\theta}_i \rangle = \sum_{j=1}^n \lambda_{ji} \alpha_j$ .

La expresión (1.29) se escribe en forma matricial como  $\boldsymbol{\beta} = \mathbf{K}\boldsymbol{\alpha}$ , donde  $\boldsymbol{\alpha}$  y  $\boldsymbol{\beta}$  son vectores con dimensión  $n$  que representan a  $\mathbf{x}$  sobre  $\mathfrak{X}$  e  $\mathbf{y}$  sobre  $\mathfrak{Y}$  por sus respectivas bases. Cabe anotar que la transformación lineal  $\mathcal{K}$  se representa completamente mediante  $\mathbf{K}$ , que es una matriz de orden  $n$  con elementos

$$\lambda_{ji} = \langle \mathcal{K} \{\varphi_i\}, \widehat{\theta}_j \rangle \quad (1.30)$$

### 1.2.3. Representación integral de sistemas lineales

En el caso de transformaciones lineales, la descripción de la relación entrada-salida en grupos de reducidos elementos se puede generalizar a todo el sistema lineal empleando la propiedad de superposición. De otra parte, la presunción de linealidad en la transformación exige que la señal a la salida sea proporcional a la entrada, pero de ninguna manera que

dependa de procesos remanentes o transitorios existentes en el sistema. En lo sucesivo, el análisis de transformaciones se realiza para sistemas invariantes en el tiempo.

**Operadores lineales en espacios integrables.** El análisis de los sistemas puede llevarse a cabo mediante cualquiera de las distintas formas de representación de las señales. En el caso discreto, se puede emplear la representación de transformaciones lineales sobre espacios de dimensión finita, correspondientes al modelo de señal dado en (1.4). Sin embargo, se tiene que la implementación práctica de los métodos de análisis es compleja, debido a que se hace necesario una gran cantidad de mediciones para los espacios de entrada en la descripción de cada transformación; sumado al hecho de que la misma cantidad de transformaciones puede ser muy alta. Por tal razón, esta forma de análisis de sistemas es empleada en casos específicos de espacios de entrada con bajo número de dimensiones y transformaciones relativamente simples.

El análisis de sistemas es preferible llevarlo a cabo, basados en la representación integral de señales, empleando operadores lineales, los cuales se definen sobre la transformación lineal acotada y determinada en espacios integrables  $L^2(T)$ . La interpretación física de esta restricción tiene que ver con el cumplimiento de la condición de estabilidad de los sistemas, que implica que a una señal de entrada con energía le debe corresponder también otra señal de energía a la salida.

Empleando el modelo de bases continuas en la representación generalizada de Fourier (1.13), las respectivas señales de entrada y salida se describen como:

$$x(t) = \int_S v_x(s) \phi(t, s) ds, \quad (1.31a)$$

$$y(t) = \int_S v_y(s) \phi(t, s) ds. \quad (1.31b)$$

Así, se define el operador  $\mathcal{K}$ , tal que

$$y(t) = \mathcal{K} \{x(t)\} = \int_S v_y(s) \psi(t, s) ds$$

donde  $\psi(t, s) = \mathcal{K} \{\phi(t, s)\}$ .

La densidad de salida, que de acuerdo con (1.12) es  $v_y(s) = \int_T y(t) \theta(s, t) dt$  y al reemplazar (1.31b), resulta en

$$v_y(s) = \int_T \int_S v_x(\sigma) \psi(t, \sigma) \theta(s, t) d\sigma dt = \int_S K(s, \sigma) v_x(\sigma) d\sigma,$$

donde el núcleo de la anterior integral, definido como,

$$K(s, \sigma) = \int_T \psi(t, \sigma) \theta(s, t) dt \quad (1.32)$$

representa la transformación  $\mathcal{K}$  y corresponde al sentido de los coeficientes (1.30), en los cuales los coeficientes  $i$  y  $j$  se convierten en las variables  $\sigma$  y  $s$  de (1.32), mientras las expresiones de representación de entrada y salida, en términos de la base  $\varphi(t, s)$ , se obtienen de las respectivas representaciones integrales. Así por ejemplo, se puede tomar la base dada en (1.21), para la cual se cumple que  $\phi(t, s) = \delta(t - s)$ , con  $T, S \subset (-\infty, \infty)$ . En este caso, de acuerdo con (1.22), se tiene que  $v_x(s) = x(s)$  y  $v_y(s) = y(s)$ . La reacción a la función base, especialmente notada como

$$\psi(t, \sigma) \triangleq h(t, s) \quad (1.33)$$

se denomina *respuesta a impulso*, que como función en el tiempo describe la reacción del operador cuando a la entrada se tiene la función  $\delta$  dada en el momento de tiempo  $s$ . Luego, de (1.32) se tiene

$$K(s, \sigma) = \int_{-\infty}^{\infty} \delta(s - t) h(t, \sigma) dt = h(s, \sigma) \quad (1.34)$$

con lo cual, la salida (1.31b) se determina como

$$y(t) = \int_T h(t, \tau) x(\tau) d\tau \quad (1.35)$$

La expresión (1.35) es la primera forma de análisis de sistemas, que corresponde al empleo en calidad de núcleo de su reacción a la función  $\delta(t)$ , conocida como el *método de la integral de superposición*. En otras palabras, para describir el comportamiento de un sistema en el dominio del tiempo, se puede probar aplicando la función elemental  $\delta(t)$  a la entrada; la salida resultante suficiente para la descripción de la respuesta del sistema.

Otra base frecuentemente empleada corresponde a  $\phi(t, s) = e^{j2\pi st}$ ,  $T, S \subset (-\infty, \infty)$ . En este caso, se obtiene que  $v_x(s) = \mathcal{F}\{x(t)\} = X(s)$ , además,  $v_y(s) = \mathcal{F}\{y(t)\} = Y(s)$ . De acuerdo con (1.34), se tiene

$$\psi(t, s) = \int_T h(t, \tau) e^{j2\pi s\tau} d\tau \triangleq H(t, s) \quad (1.36)$$

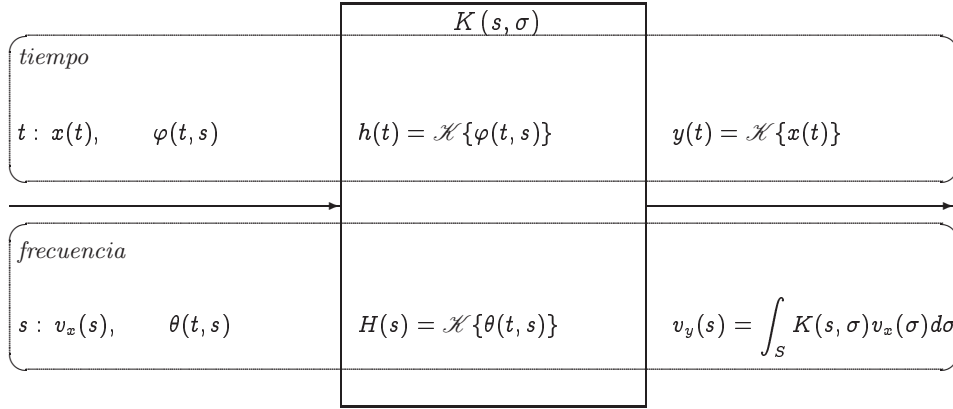
Por consiguiente, de (1.34) y (1.35) se obtiene

$$Y(f) = \int_{-\infty}^{\infty} H(f, \nu) X(\nu) d\nu \quad (1.37)$$

$$\text{donde } H(f, \nu) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h(t, \tau) e^{-j2\pi ft} e^{j2\pi \nu \tau} dt d\tau.$$

La respuesta de frecuencia  $H(f, \nu)$ , denominada *la función de transferencia*, corresponde a la segunda forma de análisis de sistemas conocida como el *método de análisis espectral*.

La relación entre la función de respuesta a impulso y la función de transferencia, así como con las respectivas bases de representación en la transformación lineal de señales se muestra en la Figura 1.3.



**Figura 1.3.** Transformación lineal e invariante de señales

Cabe anotar, que en algunos casos es importante representar las señales de entrada y salida con diferentes bases, así  $\varphi(t, s)$  puede corresponder a la entrada, mientras,  $\hat{\varphi}(t, s)$  a la salida, con lo cual la respectiva densidad de salida es:

$$v_y(s) = \int_s K(s, \sigma) v_x(\sigma) d\sigma$$

donde  $K(s, \sigma) = \int_T \hat{\theta}(s, t) \psi(t, \sigma) dt, \quad \psi(t, s) = \mathcal{K}\{\varphi(t, s)\}$

El tercer método de análisis de sistemas lineales se basa en su descripción generalizada mediante ecuaciones diferenciales, también lineales, pero de orden finito y conocido como el *método del operador diferencial de orden finito*,

$$\alpha_0(t) y(t) + \alpha_1(t) \frac{dy}{dt} + \cdots + \alpha_m(t) \frac{d^m y}{dt^m} = \beta_0(t) x(t) + \beta_1(t) \frac{dx}{dt} + \cdots + \beta_n(t) \frac{d^n x}{dt^n}, \quad (1.38)$$

siendo, respectivamente, las funciones continuas de entrada y salida  $\{\alpha_m(t)\}$  y  $\{\beta_n(t)\}$ .

**Sistemas lineales e invariantes en el tiempo.** La condición de invariabilidad, implica que el operador dependa sólo de la diferencia de los argumentos,  $y(t - t_0) = \mathcal{K}\{x(t - t_0)\}$ , que en los sistemas lineales, implica que la respuesta a impulso  $h(t, \tau)$  dependa solamente de la diferencia de los argumentos,  $h(t, \tau) = h(t - \tau)$ , con lo que (1.35) se convierte en

$$y(t) = \int_T h(t - \tau) x(\tau) d\tau \quad (1.39)$$

Así mismo, la descripción, dada en términos de la frecuencia lineal,  $f$ , para el operador

invariante en el tiempo, contiene el ajuste de los argumentos en (1.37):

$$\begin{aligned} H(f, \nu) &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h(t - \tau) e^{-j2\pi ft + j2\pi \nu \tau} dt d\tau = \int_{-\infty}^{\infty} h(\sigma) e^{-j2\pi f \sigma} \int_{-\infty}^{\infty} e^{j2\pi(f - \nu)\tau} d\tau d\sigma \\ &= H(f) \delta(f - \nu) \end{aligned}$$

por consiguiente

$$Y(f) = \int_{-\infty}^{\infty} H(f) \delta(f - \nu) X(\nu) d\nu = H(f) X(f) \quad (1.40)$$

Asumiendo  $x(t) = e^{j2\pi ft}$ , la convolución dada en (1.39) se define como:

$$\begin{aligned} y(t) &= \int_{-\infty}^{\infty} h(\tau) x(t - \tau) d\tau = \int_{-\infty}^{\infty} h(\tau) e^{j2\pi f(t - \tau)} d\tau = e^{j2\pi ft} \int_{-\infty}^{\infty} h(\tau) e^{-j2\pi f \tau} d\tau \\ &= e^{j2\pi ft} \mathcal{F}\{h(t)\} = e^{j2\pi ft} H(2\pi f) \end{aligned}$$

Luego, la función de transferencia  $H(2\pi f)$ , expresada en términos de frecuencia angular, corresponde a la TF de la función respuesta al impulso:

$$H(2\pi f) = \mathcal{F}\{h(t)\} = \int_{-\infty}^{\infty} h(t) e^{-j2\pi ft} dt$$

Por último, aplicando la TIF a (1.40), la salida se puede expresar en forma alterna:

$$y(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} H(2\pi) X(2\pi) e^{-j2\pi t} d\omega$$

El método del operador diferencial asume que en los sistemas invariantes en el tiempo los parámetros de entrada y salida en la ecuación diferencial (1.38) son constantes, con lo cual la correspondiente función respuesta a impulso de (1.39) se determina como [4]:

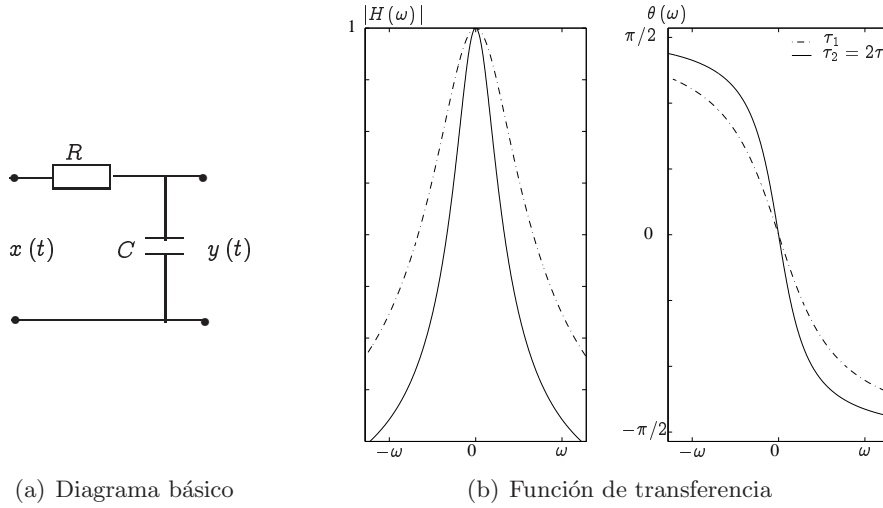
$$h(t - \tau) = \begin{cases} \sum_{k=1}^m \alpha_k e^{p_k(t - \tau)}, & t \geq \tau \\ 0, & t < \tau \end{cases} \quad (1.41)$$

donde  $p_k$  son las raíces del polinomio,  $Q(p) = \sum_{k=1}^m \alpha_k p^k = 0$ .

Al operador de (1.41) le corresponde la función de transferencia, que se expresa en la siguiente forma racional:

$$H(f) = \sum_{k=1}^m \frac{\alpha_k}{j\omega - p_k} = \frac{P(p)}{Q(p)}, \quad \alpha_k, \beta_k \in \mathbb{R}, \quad P(p) = \sum_{k=1}^n \beta_k p^k$$

**Ejemplo 1.3.** Determinar la función de transferencia del circuito  $RC$  en la Figura 1.4(a).



**Figura 1.4.** Análisis espectral del circuito  $RC$

Aplicando la relación del divisor de voltaje, se tiene:

$$y(t) = \frac{Z_c}{R + Z_c} e^{j\omega t}$$

donde la impedancia de la capacitancia es  $Z_c = 1/j\omega C$ . Así,

$$H(\omega) = \frac{(1/j\omega C)}{R + (1/j\omega C)} = \frac{1}{1 + j\tau\omega} \quad (1)$$

siendo  $\tau = RC$ . Al convertir en forma polar la expresión (1) se obtienen las respectivas *magnitud*,  $|H(\omega)| = \left(1 + (\omega\tau)^2\right)^{-1/2}$ , y *desfase*, que se muestran en la Figura 1.4(b).

#### 1.2.4. Representación de sistemas no lineales

El comportamiento de sistemas variantes en el tiempo, pero que trabajan en régimen no lineal, se describe por ecuaciones diferenciales también no lineales. La integración exacta de tales ecuaciones, en la mayoría de casos prácticos, es imposible. Actualmente, existe una serie de métodos de solución aproximada, los cuales básicamente, representan el dispositivo en análisis como una secuencia de conexiones lineales e invariantes sin retroalimentación. En este caso, se deben hacer los ajustes necesarios derivados de la aparición de nuevas componentes espectrales al paso de señales por sistemas no lineales. Aunque, cuando se tienen sistemas no lineales con lazos de retroalimentación (osciladores, sistemas de seguimiento, etc.), su análisis se hace mucho más complejo. Algunos de los métodos de representación corresponden a los descritos a continuación:

**Análisis del régimen de estado estable.** En este caso, se toma el modelo de estabilidad asintótica del sistema, aunque, de igual forma se puede asumir la variabilidad o invariabilidad de sus parámetros en el tiempo. El análisis puede realizarse tanto en el tiempo como en la frecuencia.

El método espectral se puede emplear cuando el sistema es abierto y consiste de una conexión en cascada de dispositivos lineales invariantes o no lineales variantes en el tiempo. En primera instancia, puede considerarse el modelo de un dispositivo no lineal con una función de transformación entrada-salida, descrita por  $y(t) = \mathcal{K}\{x(t)\}$ , a cuya salida le corresponde la TF:

$$Y(\omega) = \int_{-\infty}^{\infty} \mathcal{K}\{x(t)\} e^{-j\omega t} dt \quad (1.42)$$

que determina de forma analítica la función que aproxima la característica del dispositivo no lineal.

En general, la señal de entrada se puede representar mediante su descomposición espectral, por ejemplo, mediante la serie de Fourier, luego se obtiene para cada una de las componentes, su respectiva salida después de cada dispositivo, teniendo en cuenta que la señal total de salida superpone todas las respuestas elementales del sistema:

$$y = \mathcal{K}\left\{\sum_{k=1}^n x_n e^{jk\omega_0 t}\right\}$$

La segunda forma de representación de un sistema, en régimen de estado estable, corresponde a su desarrollo en el tiempo. Si se asume la invariabilidad del sistema, todas las derivadas en el tiempo de los respectivos parámetros deben ser cero, con lo cual se puede analizar el régimen de estado estable con base en la respectiva ecuación diferencial, que describe el sistema, la cual después de hacerse cero se convierte en una ecuación algebraica simple. La solución de esta clase de ecuaciones se puede realizar por métodos gráficos o de aproximaciones sucesivas numéricas [6].

**Ejemplo 1.4.** Hallar la DEE de la corriente de salida relacionada con el voltaje de entrada mediante la expresión,

$$i(t) = \begin{cases} k(u_1 e^{-\alpha t} + u_0) > 0, & t_1 \leq t \leq t_2 \\ 0, & \text{otros valores de } t \end{cases}$$

donde  $k$  es una constante.

La densidad espectral de energía del voltaje de entrada, teniendo en cuenta el valor del retardo  $t_1$  tiene la forma

$$U(\omega) = \int_{t_1}^{\infty} u_1 e^{-\alpha(t-t_1)} e^{-j\omega t} dt = u_1 \frac{e^{j\omega t_1}}{\alpha + j\omega}$$

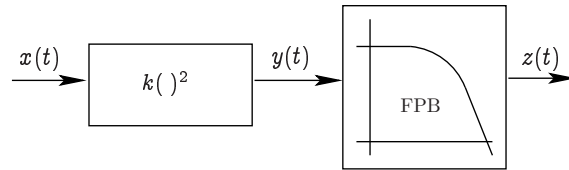
mientras a la salida, a partir de (1.42), le corresponde la densidad espectral,

$$I(\omega) = \int_{t_1}^{t_2} k \left( u_1 e^{-\alpha(t-t_1)} + u_0 \right) e^{-j\omega t} dt = k e^{-j\omega t_1} \int_0^{t_2-t_1} (u_1 e^{-\alpha t} + u_0) e^{-j\omega t} dt$$

$$= k u_1 \frac{e^{-j\omega t_1}}{\alpha + j\omega} \frac{\alpha e^{-(\alpha+j\omega)(t_2-t_1)} - (\alpha + j\omega) e^{-\alpha(t_2-t_1)} + j\omega}{j\omega}$$

cuya forma se hace compleja, como resultado de la no linealidad del proceso.

**Ejemplo 1.5.** Analizar el régimen estable de una señal compuesta de dos componentes armónicos:  $x(t) = x_1 \cos(\omega_1 t) + x_2 \cos(\omega_2 t)$ , a su paso el sistema mostrado en la Figura 1.5, compuesto por un elevador al cuadrado (con función entrada-salida  $y(t) = kx^2(t)$ ) y un filtro  $RC$  pasabajos, como el descrito por la función de transferencia (1) del ejemplo 1.3.



**Figura 1.5.** Detector cuadrático

A la salida del elevador al cuadrado se tiene

$$\begin{aligned} y(t) &= kx^2(t) = k(x_1 \cos(\omega_1 t) + x_2 \cos(\omega_2 t))^2 \\ &= kx_1^2 \cos^2(\omega_1 t) + 2kx_1 x_2 \cos(\omega_1 t) \cos(\omega_2 t) + kx_2^2 \cos^2(\omega_2 t) \\ &= \frac{k}{2} (x_1^2 + x_2^2) + \frac{k}{2} x_1^2 \cos(2\omega_1 t) + \frac{k}{2} x_2^2 \cos(2\omega_2 t) + kx_1 x_2 (\cos(\omega_3 t) + \cos(\omega_4 t)) \end{aligned}$$

donde  $\omega_3 = \omega_1 - \omega_2$  y  $\omega_4 = \omega_1 + \omega_2$ .

Teniendo en cuenta las expresiones obtenidas para la magnitud y desfase del filtro  $RC$  en el ejemplo 1.3, la salida del detector cuadrático tiene la forma,

$$\begin{aligned} z(t) &= \frac{k}{2} (x_1^2 + x_2^2) + \frac{k}{2} \frac{x_1^2}{\sqrt{1 + 4\omega_1^2 \tau^2}} \cos(2\omega_1 t + \theta_1) + \frac{k}{2} \frac{x_2^2}{\sqrt{1 + 4\omega_2^2 \tau^2}} \cos(2\omega_2 t + \theta_2) + \\ &+ k \frac{x_1 x_2}{\sqrt{1 + \omega_3^2 \tau^2}} \cos(\omega_3 t + \theta_3) + k \frac{x_1 x_2}{\sqrt{1 + \omega_4^2 \tau^2}} \cos(\omega_4 t + \theta_4) \end{aligned}$$

donde  $\theta_{1,2} = -\arctan(2\omega_{1,2}\tau)$ ,  $\theta_{3,4} = -\arctan(\omega_1 \pm \omega_2\tau)$ . Si se asume un valor suficientemente grande la constante  $\tau$ , la salida del filtro pasabajos (FPB) se puede aproximar al siguiente valor:

$$z(t) \approx \frac{k}{2} (x_1^2 + x_2^2)$$

**Método de series funcionales.** También conocido como el *método de Wiener* [7], puede ser catalogado como heurístico y se fundamenta en la posibilidad de representar sistemas no lineales e invariantes en el tiempo en forma de series.



Sea un sistema con entrada  $x(t)$  y salida  $y(t)$ ; ambas relacionadas por el operador  $\mathcal{K}$ :

$$y(t) = \mathcal{K}\{x(t)\}$$

En general, el operador  $\mathcal{K}$  depende del carácter del sistema, su objeto de análisis y tipo de entradas. En el caso concreto, se escoge una clase de sistemas con memoria finita, asumiendo que el operador es biyectivo. Esta restricción no aplica en sistemas en los cuales existan puntos de equilibrio dependientes de las condiciones iniciales de entrada (sistemas con memoria infinita).

Una forma de representación de la no linealidad en un sistema, más angosta que las correspondientes ecuaciones diferenciales, corresponde a las *series de Volterra* [8]:

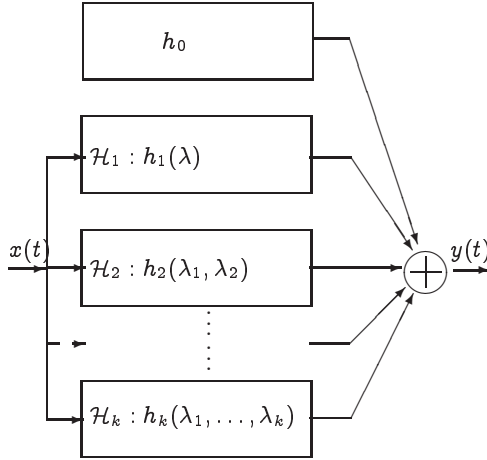
$$y(t) = h_0(t) + \sum_k \int \cdots \int h_k(t, \tau_1, \dots, \tau_k) x(\lambda_1) \cdots x(\lambda_k) d\lambda_1 \cdots d\lambda_k \quad (1.43)$$

En particular, cuando se tienen ecuaciones diferenciales de primer orden [9],

$$\frac{dy}{dt} = f(t, y(t)) + g(t, y(t), x(t))$$

la representación (1.43) es viable, por ejemplo, cuando la función  $f(t, y(t))$  con respecto a  $y$  se puede aproximar por la serie de Taylor, mientras la función  $g$  no dependa de  $y(t)$ .

Con el objeto de simplificar el análisis, el modelo de representación se puede limitar hasta los sistemas invariantes en el tiempo, cuando se asume que es cierta la propiedad  $\mathcal{K}\{x(t + \tau)\} = y(t + \tau)$ ,  $\forall \tau$ , por esta razón, la serie funcional de Volterra (1.43) toma una forma más simple,



**Figura 1.6.** Representación por series de Volterra

$$y(t) = h_0 + \int_{-\infty}^{\infty} h_1(\lambda_1) x(t - \lambda_1) d\lambda_1 + \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h_2(\lambda_1, \lambda_2) x(t - \lambda_1) x(t - \lambda_2) d\lambda_1 d\lambda_2 + \cdots \quad (1.44)$$

Las funciones  $h_k(\lambda_1, \dots, \lambda_k)$  son denominadas los *núcleos de Volterra*, que al igual que en los sistemas lineales, representan la respuesta a la función  $\delta(t)$ . El primer término  $h_0$  se puede tomar como componente constante de la representación, o bien, como las condiciones iniciales, por cuanto  $y(t) = h_0$ , cuando  $x(t) = 0$ . En cuanto al segundo término,

$\int h_1(\lambda_1) x(t - \lambda_1) d\lambda_1$ , si cumple la condición de realización física,  $h_1(t) = 0, \forall t < 0$ , entonces éste corresponde a la parte lineal del sistema.

El tercer término,  $\int \int h_2(\lambda_1, \lambda_2) x(t - \lambda_1) x(t - \lambda_2) d\lambda_1 d\lambda_2$ , corresponde a la convolución doble de la señal de entrada  $x(t)$  con la respuesta a impulso  $h_2(\lambda_1, \lambda_2)$ . Si se asume la independencia de los términos  $h_2(\lambda_1, \lambda_2) = h(\lambda_1)h(\lambda_2)$ , entonces el tercer término es:

$$\begin{aligned} \int \int h_2(\lambda_1, \lambda_2) x(t - \lambda_1) x(t - \lambda_2) d\lambda_1 d\lambda_2 \\ = \int h(\lambda_1) x(t - \lambda_1) d\lambda_1 \int h(\lambda_2) x(t - \lambda_2) d\lambda_2 = y(t) y(t) = y^2(t) \end{aligned}$$

esto es, el tercer término puede ser entendido como la descripción de un sistema cuadrático.

En general, asumiendo que todos los núcleos de Volterra cumplen las condiciones:

- *Realización física.*  $h_k(\lambda_1, \dots, \lambda_n) = 0, \forall \tau_k < 0, k = 1, \dots, n$
- *Simetría.*  $h_k(\lambda_1, \dots, \lambda_n) = h_k(\lambda_n, \dots, \lambda_1) = \dots$ , esto es, el núcleo es el mismo cualquiera que sea el orden de sus argumentos.

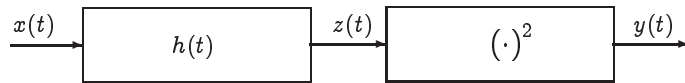
entonces, un sistema no lineal dado con memoria finita puede ser representado por la serie de Volterra (1.44), en la cual cada término puede ser asociado con su respectivo operador elemental  $\mathcal{H}_k$ , como se muestra en la Figura 1.6:

$$y(t) = h_0 + \sum_{k \geq 0} \mathcal{H}_k \{x(t)\} \quad (1.45)$$

donde  $\mathcal{H}_k \{x(t)\} = \int \dots \int h_k(\lambda_1, \dots, \lambda_k) x(\lambda_1) \dots x(\lambda_k) d\lambda_1 \dots d\lambda_k$ .

---

**Ejemplo 1.6.** Analizar la convergencia del sistema invariable en el tiempo, descrito por la función linealidad  $y(t) = z(t) (1 + z^2(t))^{-1}$ , representado en la Figura 1.7.



**Figura 1.7.** Sistema cuadrático simple

La descomposición de la función, mediante análisis en la serie de Taylor, tiene la forma:

$$y(t) = \sum_{k=0}^{\infty} (-1)^k (z(t))^{2k+1} \quad (1)$$

Reemplazando, por el núcleo básico de primer orden  $z(t) = h(t) * x(t)$ , se obtiene la respectiva serie de Volterra

$$y(t) = \sum_{k=0}^{\infty} (-1)^k \left( \int_{-\infty}^{\infty} h(\lambda) x(t - \lambda) d\lambda \right)^{2k+1} = \sum_{k=0}^{\infty} \mathcal{H}_{2k+1} \{x(t)\} \quad (2)$$

donde los núcleos son de la forma  $h_{2k+1}(\lambda_1, \dots, \lambda_{2k+1}) = (-1)^k h(\lambda_1) \cdots h(\lambda_{2k+1})$ .

Por cuanto la serie de Taylor (1), obtenida para el sistema cuadrático simple, converge solo para valores de  $z(t) < 1$ , entonces la respectiva serie (2) converge en aquellos momentos del tiempo cuando  $|z(t)| \geq 1$ . Lo anterior implica, que en el ejemplo del sistema cuadrático simple, la serie de Volterra se puede emplear para señales de entrada, cuya salida  $z(t)$  no exceda el valor de 1.

---

Cabe anotar que la convergencia de la serie de Taylor se exige para todo el dominio de representación  $t$ . Así por ejemplo, si en la Figura 1.7, en vez del elevador a cuadrado, se tiene un limitador bilateral del tipo  $y(t) = a \operatorname{sgn}(z(t))$ , entonces para esta función con rompimiento en el punto  $t = 0$ , la serie de Taylor no existe en ese mismo punto, y por lo tanto, tampoco existe la representación en series de Volterra.

## Problemas

**Problema 1.7.** Demostrar que el siguiente conjunto de señales es ortogonal y calcular su TF directa:

$$\Phi_n(t) = \text{sinc } \Omega(t - nT), \quad n = 0, \pm 1, \pm 2, \dots, -\infty < t < \infty.$$

**Problema 1.8.** Calcular los coeficientes de la serie de Fourier y la TF de la función dada en la forma,  $x(t) = \sum_{n=-\infty}^{\infty} \delta(t - nT)$ .

**Problema 1.9.** Hallar la TF de un pulso Gaussiano  $x(t) = a \exp -\pi(t/\tau)^2$  y verificar el efecto de extensión recíproca de escala para  $\tau = 1$  y  $\tau = 2$ . Calcular la diferencia de las DEE en los puntos de intersección.

**Problema 1.10.** Las señales  $x_p(t) = x_p(-t)$  y  $x_i(t) = -x_i(-t)$  se relacionan con la función  $y(t)$  como:

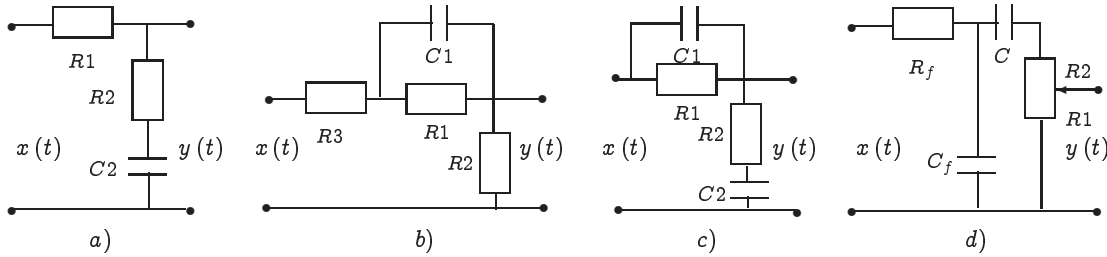
$$x_p(t) = y(t) + y(-t)$$

$$x_i(t) = y(t) - y(-t)$$

Hallar la relación entre las respectivas DEE  $X_k(\omega) = \mathcal{F}\{x_k(t)\}$ ,  $k \in \{p, i\}$  con la densidad  $Y(\omega)$ .

**Problema 1.11.** Sea  $X(\omega) = \mathcal{F}\{x(t)\}$ . Hallar la señal  $y(t) = \mathcal{F}^{-1}\{Y(\omega)\}$ , para cada uno de los siguientes casos: a)  $Y(\omega) = X^2(\omega)$ , b)  $Y(\omega) = X(\omega)X^*(\omega)$  y c)  $Y(\omega) = X^*(\omega)$ .

**Problema 1.12.** Conformar las ecuaciones diferenciales de los circuitos dados en la Figura 1.8.



**Figura 1.8.** Circuitos de filtración

**Problema 1.13.** Determinar la respuesta a impulso para cada uno de los circuitos descritos por las siguientes ecuaciones diferenciales:

$$\begin{aligned} T \frac{dy}{dt} - y &= kx & y &= k \left( x + T \frac{dx}{dt} \right) \\ y &= k \left( x + 2k_1 T \frac{dx}{dt} + T^2 \frac{d^2x}{dt^2} \right) & T^2 \frac{d^2y}{dt^2} + 2k_1 T \frac{dy}{dt} + y &= kx \\ T^2 \frac{d^2y}{dt^2} - 2k_1 T \frac{dy}{dt} + y &= kx & \frac{1}{\omega^2} \frac{d^2y}{dt^2} + y &= kx \end{aligned}$$

**Problema 1.14.** Comparar los espectros de entrada y salida del dispositivo no lineal con tensión de salida  $i = a_0 + a_1(u - u_0) + a_2(u - u_0)^2 + a_3(u - u_0)^3$ , cuando el voltaje de entrada tiene la forma  $u = u_1(t) + u_0$ , asumiendo que la componente variable corresponde a un pulso triangular,

$$u_1(t) = \begin{cases} bt, & 0 \leq t \leq T \\ 0 & \text{otros valores de } t \end{cases}$$

## 1.3. Discretización de señales y sistemas

### 1.3.1. Discretización uniforme

El empleo de sistemas discretos en campos donde las señales son continuas supone su previa acomodación, esto es, la *discretización* de las señales. El modelo conveniente para la discretización ideal, o muestreo de una señal continua  $x_c(t)$ , está dado por su multiplicación con alguna señal periódica de discretización  $x_d(t)$ , con el fin de obtener la señal discretizada  $x(k\Delta t)$ :

$$x(k\Delta t) = x_c(t)x_d(t) = \sum_k x_c(k\Delta t)\delta(t - k\Delta t) \quad (1.46)$$

donde  $x_d(t) = \sum_{k=-\infty}^{\infty} \delta(t - k\Delta t)$ , siendo  $\Delta t$  el período de discretización.

A partir de la expresión (1.46) se observa que la señal discretizada  $x(t)$  se define solamente en los momentos equidistantes de tiempo  $t = k\Delta t$ , luego entonces, parte de la información de la señal original  $x_c(t)$  se pierde durante su discretización.

**Teorema 1.3.** (*Discretización de Kotelnikov.*) Sea una señal continua  $x_c(t)$ , para la cual no se consideran componentes espectrales mayores a  $\omega_{\max} = 2\pi f_{\max}[\text{rad/s}]$ , donde  $f_{\max}$  es el máximo valor de frecuencia para la cual  $X(\omega) \neq 0$ , entonces, toda la información de la señal continua estará enteramente contenida en los valores  $x_c(n\Delta t)$ , asegurando que se cumpla la desigualdad:

$$\Delta t \leq 1/2f_{\max} \quad (1.47)$$

El teorema 1.3 permite representar la señal continua  $x_c(t)$  en forma de la serie:

$$x_c(t) = \sum_{k=-\infty}^{\infty} x_c(k\Delta t) \frac{\sin(\omega_{\max}(t - k\Delta t))}{\omega_{\max}(t - k\Delta t)} \quad (1.48)$$

Al comparar la serie (1.48) con la representación (1.4), se observa que las funciones base de descomposición son las siguientes:

$$\phi_k(t) = \frac{\sin(\omega_{\max}(t - k\Delta t))}{\omega_{\max}(t - k\Delta t)}$$

La serie de Kotelnikov (1.48) permite el restablecimiento de la función inicial  $x_c(t)$ , en cualquier momento del tiempo  $t$  (dentro del intervalo de análisis). Sin embargo, el restablecimiento de  $x_c(t)$  exige realizar la sumatoria sobre una cantidad infinita de términos, lo cual prácticamente es imposible. De otra manera, la implementación del teorema de discretización conlleva inevitablemente a errores de representación.

Cabe anotar, que los valores sucesivos de los tiempos discretización uniforme, descritos por la malla  $\{n\Delta t : n \in \mathbb{Z}\}$ , están separados a un intervalo constante e igual a  $\Delta t$ , cumpliendo (1.47), por lo que en la práctica es común normalizar su valor, esto es,

$$x_c(n\Delta t) = x_d[n]$$

### 1.3.2. Transformadas ortogonales discretas

La realización de cualquier forma de representación integral de señales mediante sistemas reales de proceso digital, dada una señal discreta  $x[n]$ , se encuentra con algunas dificultades insuperables. En particular, se necesitan procesadores con memoria infinita para guardar, tanto la infinita cantidad de valores de la serie de entrada  $x[n]$ , como el conjunto de valores del continuo de la respectiva representación espectral  $H(e^{j\Omega})$ . En consecuencia, empleando el modelo de las respectivas bases continuas, en la representación generalizada de Fourier en (1.12), y para efectos de realización práctica, el conjunto de transformadas ortogonales discretas se define estrictamente sobre sucesiones discretas,  $x[n]$ , con longitud finita  $N$ , o bien sobre sucesiones periódicas con periodo  $N$ , de la siguiente forma:

$$x[k] = \sum_{n=0}^{N-1} v[n] \phi_k[n], \quad k = 0, 1, \dots, N-1 \quad (1.49a)$$

$$v[n] = \frac{1}{N} \sum_{k=0}^{N-1} x[k] \phi_n^*[k], \quad n = 0, 1, \dots, N-1 \quad (1.49b)$$

Entre los diferentes conjuntos ortogonales analizados en la sección §1.1.3, en el proceso digital de señales moderno, encuentran aplicación las transformadas discretas de Fourier (con sus respectivos casos particulares de la T. *Hartley* y la T. *discreta de cosenos*), de *Walsh* y *Haar*, entre otras. En el caso de la Transformada discreta de Fourier (TDF), en calidad de sistema base se emplea la base ortogonal exponencial (1.8), pero en su forma discretizada:

$$\phi_k^*(n) = e^{jkn2\pi/N} \triangleq W_N^{kn}, \quad n = 0, 1, \dots, N-1 \quad (1.50)$$

El *factor de pivote*  $W_N = e^{j2\pi/N}$  cumple la condición de ortogonalidad:

$$\sum_{n=0}^{N-1} W_N^{kn} W_N^{-mn} = \begin{cases} N, & (k-m) = lN, l \in \mathbb{N} \\ 0, & (k-m) \neq lN \end{cases} \quad (1.51)$$

Así mismo, este factor es periódico con período  $N$ , tanto en el sentido de la variable de frecuencia  $n$ , como el de la variable de tiempo  $k$ :

$$W_N^{kn} = W_N^{(k+N)n} = W_N^{k(n+N)} \quad (1.52)$$

Teniendo en cuenta las ecuaciones (1.50), (1.51) y (1.52), se determina el par de transformadas discretas de Fourier, en función del factor de pivote, de la siguiente forma:

$$x[k] = \sum_{n=0}^{N-1} X[n] W_N^{-kn}, \quad k = 0, 1, \dots, N-1 \quad (1.53a)$$

$$X[n] = \frac{1}{N} \sum_{k=0}^{N-1} x[k] W_N^{kn}, \quad n = 0, 1, \dots, N-1 \quad (1.53b)$$

El sistema de transformadas (1.53a) y (1.53b) establecen la relación mutua e unívoca entre los valores de la serie de entrada  $x[k]$  y su representación discreta espectral  $X[n]$ . Por cuanto el factor de pivote es periódico, entonces el espectro discreto de Fourier de la señal también es periódico, esto es,  $X(n) = X(n + mN)$ ,  $m = \pm 1, \pm 2, \dots$

### 1.3.3. Representación ortogonal de sistemas en tiempo discreto

**Respuesta discreta a impulso.** Los sistemas lineales en tiempo discreto se describen por medio de ecuaciones de diferencias con parámetros constantes:

$$\sum_{n=0}^N a_n y[k - n] = \sum_{n=0}^M b_n x[k - n] \quad (1.54)$$

donde  $a_0 = 1$ , los términos  $y[k - n]$  y  $x[k - n]$  corresponden a los valores de las señales discretizadas de salida y entrada, respectivamente. Si se cumple que  $a_n \neq 0, \forall n \neq 0$ , entonces el sistema se denomina *recursivo*. La señal a la salida de tal sistema depende no solamente de las señales de entrada, sino de valores de la señal de salida en momentos anteriores. En cambio, cuando se tiene que los coeficientes  $a_n = 0, n = 1, 2, \dots, N$ , entonces, se habla de un sistema *no recursivo*, cuya señal de salida se determina solamente por la acción de entrada hasta el momento actual de tiempo.

La respuesta a impulso de un sistema discreto  $h[n]$  se define como la salida cuando la entrada corresponde a la función  $x[n] = \delta[n]$ . La señal a la salida de un sistema discreto lineal e invariante en el tiempo se define como:

$$y[n] = \sum_{k=-\infty}^{\infty} h[n, k] x[k] = \sum_{k=-\infty}^{\infty} h[n - k] x[k], \quad (1.55)$$

Asumiendo la condición de realización física del sistema, que implica  $h[n, k] = 0, \forall k > n$ , entonces, se obtiene que,  $y[n] = \sum_{k=0}^{\infty} h[k] x[n - k], n = 0, 1, \dots$

En los filtros no recursivos, la respuesta a impulso contiene una cantidad finita de términos, que no excede el valor de  $M + 1$ , donde  $M$  es el límite superior de la suma en (1.55), correspondiente a los elementos de la entrada; estos sistemas se denominan de *respuesta a impulso finita*:

$$h_{RIF}[n] = \sum_{i=n_0}^N \alpha_i \delta[n - i], \quad \forall n_0, N < \infty \quad (1.56)$$

Por el contrario, los sistemas discretos recursivos que poseen una respuesta a impulso infinita se denominan de *respuesta a impulso infinita*:

$$h_{RII} = \sum_{i=n_0}^N \alpha_i \delta[n - i], \quad |N - n_0| \rightarrow \infty \quad (1.57)$$

**Función de transferencia discreta.** Como en caso continuo, en los sistemas discretos lineales e invariantes en el tiempo con respuesta a impulso dada  $h[n]$ , se determina la señal de salida  $y[n]$ , asumiendo que se tiene como señal de entrada  $x[n] = \exp[jn\Omega]$ , esto es,

$$y[n] = \sum_{i=-\infty}^{\infty} h[i]x[n-i] = \sum_{i=-\infty}^{\infty} h[i]e^{jn\Omega} = e^{jn\Omega} \sum_{i=-\infty}^{\infty} h[i]e^{-ji\Omega}$$

teniendo en cuenta (1.36), entonces  $y[n] = e^{jn\Omega} H(e^{j\Omega})$ , donde  $H(e^{j\Omega})$  es la respuesta de frecuencia o *función de transferencia discreta*, que en forma general, puede ser compleja como se demuestra en el caso de los sistemas continuos, así,

$$Y(e^{j\Omega}) = X(e^{j\Omega}) H(e^{j\Omega}),$$

entonces,

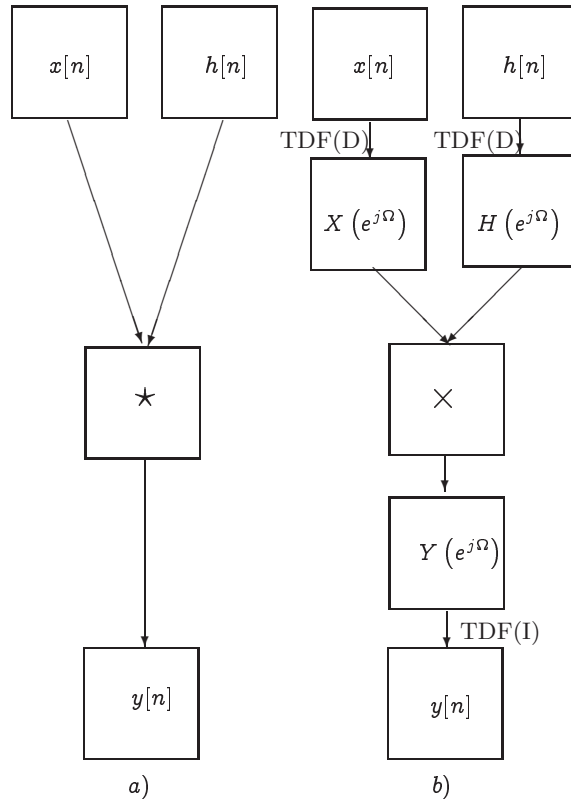
$$y[n] = x[n] * h[n]$$

Los métodos de análisis de sistemas discretos, básicamente, corresponden a los mismos métodos del caso continuo, los cuales se representan en la Figura 1.9.

De acuerdo con el desarrollo matemático anteriormente descrito de la función de transferencia, para su determinación existen varios métodos. El primer método emplea la función de impulso unitario  $\delta[n]$  como señal de entrada, un segundo método emplea la función exponencial,  $e^{jn\Omega}$ , en correspondencia con la misma definición y, por último, sin especificar la señal de entrada se puede analizar la ecuación de diferencias que concretamente describe el sistema [10].

De otra parte, el método matricial es común en la descripción de sistemas discretos, cuando el operador  $\mathcal{H}$  se representa en forma de matriz. Sean, en calidad de funciones propias de un operador, las siguientes funciones discretas  $\phi_m[k]$ ,  $m, k = 0, 1, \dots, N-1$ , esto es,

$$\mathcal{H}\{\phi_m[k]\} = H[m]\phi_m[k],$$



**Figura 1.9.** Análisis de sistemas discretos

La filtración discreta se puede entender como la multiplicación de los coeficientes  $v[n]$ ,



obtenidos de la representación de la señal de entrada con la base  $\phi_m[k]$ , por los respectivos coeficientes de la función de transferencia  $H[m]$ . La multiplicación corresponde a los coeficientes espectrales de la señal de entrada.

Las restricciones de implementación práctica hacen que en el proceso de señales se utilicen solamente bases discretas del tipo ortogonal, para los cuales la señal de salida puede ser representada por la siguiente expresión matricial:

$$\mathbf{Y}_{N \times 1} = \frac{1}{N} \mathbf{\Phi}_N^H \mathbf{H}_N \mathbf{\Phi}_N \mathbf{X}_N$$

donde  $\mathbf{Y}$  y  $\mathbf{X}$  se forman de las respectivas señales discretizadas de salida y entrada,  $\mathbf{\Phi}$  es una matriz ortogonal, con filas determinadas por las funciones discretas ortogonales  $\phi_m[k]$ ,  $m, k = 0, \dots, N - 1$ , donde  $N$  es el conjunto de la base y  $\mathbf{H}$  es la matriz ortogonal compuesta por los coeficientes  $H[m]$ . De lo anterior, se deduce que la transformación lineal discreta se representa por la matriz

$$\mathcal{H} = \frac{1}{N} \mathbf{\Phi}^H \mathbf{H} \mathbf{\Phi} \quad (1.58)$$

## Problemas

**Problema 1.15.** Sea la señal de video  $x(t) = x_0 \exp(-\alpha t) u(t)$ . Seleccionar el valor del intervalo de muestreo  $\Delta$ , de tal manera que la magnitud de la DEE en la frecuencia límite  $\omega_{\max}$  se disminuye hasta el valor de  $0.01X(0)$ .

**Ejercicio en el CP 1.2.** Sea una señal seno  $x(t)$  definida en el intervalo de tiempo  $[0, 1]$ . Hallar su representación discretizada a una velocidad de muestreo 0.01.

**Ejercicio en el CP 1.3.** Dado el siguiente código, analizar las siguientes propiedades del factor de pivote:

- periodicidad, (1.52),
- antisimetría

```
N=2,3,4,5,6,7,8; % Escala de tiempo
for n = 1:5*N
    m(n)=exp(-j*2*pi*n/N);
end
m % valores del factor de pivote
```

**Problema 1.16.** Calcular la convolución lineal de las siguientes sucesiones finitas:

$$\mathbf{x} = [2 \ -2 \ 1]^T, \mathbf{h} = [1 \ 2]^T. \quad \mathbf{x} = [2 \ -2 \ 1]^T, \mathbf{h} = [1 \ 2 \ 0 \ 0]^T.$$

**Problema 1.17.** Dado un filtro de respuesta a impulso finita, para un valor de  $N = 9$ , que se aproxima a un FPB ideal con frecuencia de corte  $\Omega_c = 0.2\pi$  y respuesta a impulso deseada  $h[n] = \sin(0.2\pi n)/(\pi n)$ . Hallar la respectiva función de transferencia para los casos de ventanas: a) rectangular, b) Hamming.

**Problema 1.18.** Sea un circuito diferenciador ideal (transformador de Hilbert), para un valor de  $N = 15$ , con respuesta a impulso deseada

$$h[n] = \begin{cases} 0, & n \text{ par} \\ \frac{2}{n\pi}, & n \text{ impar} \end{cases}$$

Hallar la respectiva función de transferencia para los siguientes casos de ventana:

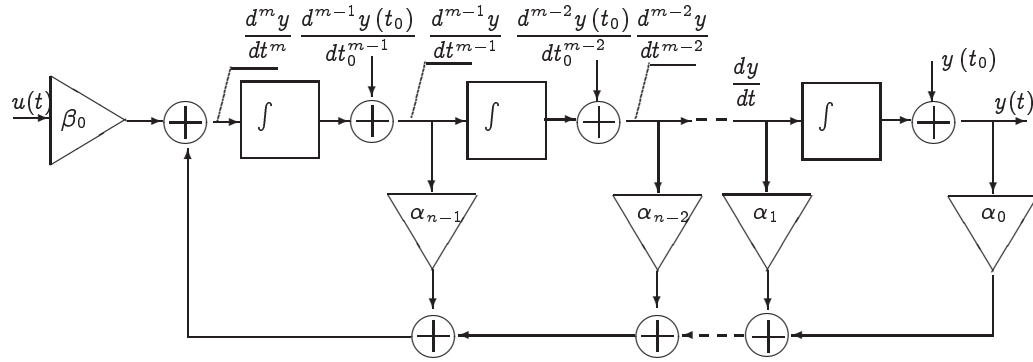
a) rectangular, b) Hamming.

**Problema 1.19.** Hallar las funciones de transferencia de los circuitos, descritos en los problemas 1.12 y 1.13.

## 1.4. Representación dinámica de sistemas

### 1.4.1. Sistemas lineales en variables de estado

A partir del método de ecuaciones diferenciales (1.38), que describe un sistema lineal con una entrada y una salida, en el numeral §1.2.3 se analiza la función de respuesta a impulso  $h(t)$  como forma de descripción del sistema. Sin embargo, el mismo método de ecuaciones diferenciales puede emplearse directamente en la representación de los sistemas dinámicos lineales. En este segundo caso, se deben tener en cuenta las siguientes condiciones:



**Figura 1.10.** Modelo de un sistema lineal de orden  $m$ .

- Las condiciones iniciales del sistema deben ser conocidas para determinar su respuesta en cualquier intervalo de tiempo  $t_0 \leq t \leq t_1$ . Además, deben estar en un número tal que describan completamente la influencia de la señal de entrada, hasta el instante  $t_0$ , sobre la señal de salida después del momento  $t_0$ .

El *estado del sistema* se define como la cantidad mínima de elementos de referencia del sistema necesaria para la descripción completa de su señal de salida, mientras las mismas variables de referencia se denominan *variables de estado*. En la práctica, el estado del sistema se describe por un vector de orden finito y, por lo tanto, se habla de un *sistema dinámico con dimensión finita*.

- La solución de la ecuación diferencial se puede implementar mediante los siguientes dispositivos de operación básica: integradores, sumadores, dispositivos no lineales, amplificadores de ganancia variable, etc. La conexión de estos dispositivos de operaciones básicas se considera el *modelo del sistema dinámico*.

La forma generalizada de una ecuación diferencial de orden  $m$ ,

$$\beta_0 u(t) = \alpha_0 y(t) + \alpha_1 \frac{dy}{dt} + \dots + \alpha_m \frac{d^m y}{dt^m}, \quad (1.59)$$

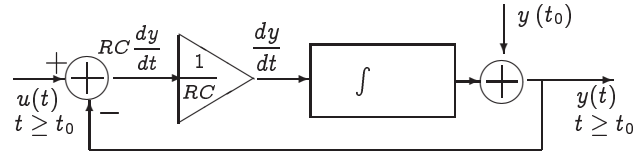
implica la selección de  $y(t), \dots, d^{m-1}y/dt^{m-1}$  en calidad de variables de estado, cuya

su solución se puede modelar empleando un procesador analógico equivalente como el mostrado en la Figura 1.10.

**Ejemplo 1.7.** Sea el circuito  $RC$  representado en la Figura 1.3, en el cual la señales de salida  $y(t)$  y entrada  $u(t)$  están relacionadas por la ecuación diferencial

$$RC \frac{dy}{dt} + y(t) = u(t) \quad (1.60)$$

El cálculo de la señal de salida (voltaje) en el intervalo  $t \geq t_0$ , implica el conocimiento de la señal de entrada para  $t \geq t_0$ , además del voltaje inicial  $y(t_0)$ , presente en el condensador en el momento  $t_0$ . De esta manera, es suficiente introducir una sola variable de estado, que es la que corresponde al voltaje  $y(t)$ . Así mismo, la expresión (1.60) es la ecuación de estado del circuito  $RC$  en análisis, en concordancia con la cual, se representa el respectivo modelo equivalente en la Figura 1.11.



**Figura 1.11.** Modelo equivalente del circuito  $RC$

Un sistema, dado por la ecuación (1.59) de orden  $m$ , puede ser descrito por ecuaciones diferenciales de primer orden. Frecuentemente, tal sistema de ecuaciones diferenciales, que es denominado *ecuación vectorial diferencial de primer orden*, es más fácil de analizar que la ecuación inicial de orden  $m$ . Así por ejemplo, en vez de la ecuación (1.59) se puede analizar el siguiente sistema:

$$\begin{aligned} x_1(t) &= y(t) \\ x_2(t) &= \frac{dy}{dt} = \frac{dx_1}{dt} \\ x_3(t) &= \frac{d^2y}{dt^2} = \frac{dx_2}{dt} \\ &\dots = \dots \\ x_n(t) &= \frac{d^{m-1}y}{dt^{m-1}} = \frac{dx_{m-1}}{dt} \end{aligned}$$

por lo tanto, se tiene que

$$\frac{du^m}{dt} = \frac{d^m y}{dt^m} = \beta_0 u(t) - \sum_{k=1}^m \alpha_{k-1} \frac{d^{k-1} y}{dt^{k-1}} = \beta_0 u(t) - \sum_{k=1}^m \alpha_{k-1} x_k(t), \quad t \geq t_0 \quad (1.61)$$

Si se introduce el vector columna de funciones  $\mathbf{x}(t) \in \{x_k : k = 1, \dots, m\}$ , para  $t \geq t_0$ , entonces en vez de la ecuación (1.59) de orden  $m$ , se analiza la siguiente ecuación equivalente

de primer orden y con dimensión  $m$ ,

$$\frac{d\mathbf{x}}{dt} = \mathbf{A}\mathbf{x}(t) + \mathbf{B}u(t) \quad (1.62)$$

donde las matrices  $\mathbf{A}_{m \times m}$  y  $\mathbf{B}_{n \times 1}$  tienen la forma

$$\mathbf{A} = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & \cdots & 1 \\ -\alpha_m & -\alpha_{m-1} & -\alpha_{m-2} & \cdots & -\alpha_1 \end{bmatrix}, \quad \mathbf{B} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ \beta_0 \end{bmatrix} \quad (1.63)$$

cuya solución requiere el vector de condiciones iniciales  $\mathbf{x}(t_0)$ . El vector  $\mathbf{x}(t)$  se denomina *vector de estado del sistema*, mientras la expresión (1.62) corresponde a la *ecuación de estado del sistema*. La señal de salida  $y(t)$  del sistema lineal en análisis se determina por medio del vector de estado en la forma,

$$y(t) = \mathbf{C}\mathbf{x}(t) \quad (1.64)$$

donde la matriz  $\mathbf{C}_{1 \times m}$  se define como,  $\mathbf{C} = [1 \ 0 \ \cdots \ 0]$ .

La generalización de ecuación diferencial (1.59), para los sistemas lineales, corresponde a la ecuación (1.38) e incluye la diferenciación de la señal  $x(t)$ , la cual debe evitarse en el análisis de sistemas estocásticos reales, entre otras razones porque es frecuente en calidad de señal de entrada tomar el ruido blanco, cuyas derivadas en el tiempo no existen.

En este sentido, asumiendo la invariabilidad de los respectivos coeficientes, la ecuación (1.38) se puede representar de forma alterna,

$$\frac{d^m y}{dt^m} + \alpha_{m-1} \frac{d^{m-1} y}{dt^{m-1}} - \beta_{m-1} \frac{du}{dt^{m-1}} + \cdots + \alpha_1 \frac{dy}{dt} - \beta_1 \frac{du}{dt} = \beta_0 u(t) - \alpha_0 y(t), \quad t \geq t_0 \quad (1.65)$$

La integración de ambas partes de la igualdad (1.65), con condiciones iniciales cero, da como resultado,

$$\frac{d^{m-1} y}{dt^{m-1}} + \alpha_{m-1} \frac{d^{m-2} y}{dt^{m-2}} - \beta_{m-2} \frac{d^{m-2} u}{dt^{m-2}} + \cdots + \alpha_1 y(t) - \beta_1 u(t) = x_m(t), \quad t \geq t_0 \quad (1.66)$$

donde

$$x_m(t) = \int_{t_0}^t (-\alpha_0 y(\tau) + \beta_0 u(\tau)) d\tau \quad (1.67)$$

De la misma manera, la ecuación (1.66) se puede escribir en la forma

$$\frac{d^{m-1} y}{dt^{m-1}} + \alpha_{m-1} \frac{d^{m-2} y}{dt^{m-2}} - \beta_{m-1} \frac{d^{m-2} u}{dt^{m-2}} + \cdots + \alpha_2 \frac{dy}{dt} - \beta_2 \frac{du}{dt} = x_m(t) - \alpha_1 y(t) - \beta_1 u(t), \quad t \geq t_0$$

cuya integración conlleva a la ecuación,

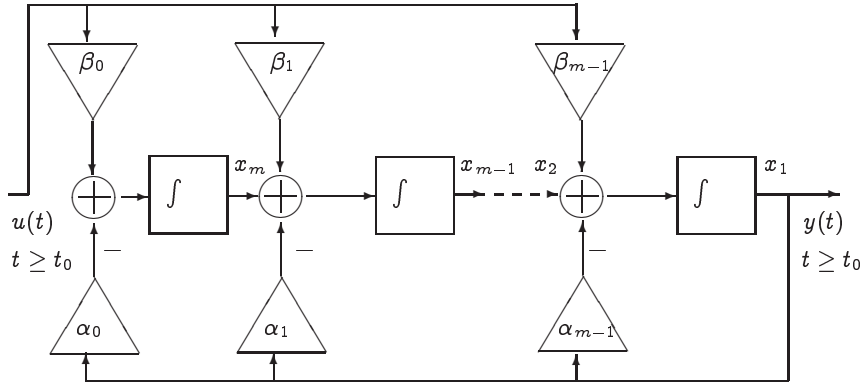
$$\frac{d^{m-2}y}{dt^{m-2}} + \alpha_{m-1} \frac{d^{m-3}y}{dt^{m-3}} - \beta_{m-1} \frac{d^{m-3}u}{dt^{m-3}} + \cdots + \alpha_2 y(t) - \beta_1 u(t) = x_{m-1}(t), \quad t \geq t_0$$

donde

$$x_{m-1}(t) = \int_{t_0}^t (x_m(t) - \alpha_1 y(\tau) + \beta_1 u(\tau)) d\tau \quad (1.68)$$

La sucesión descrita de operaciones, repetida  $m$  veces, da como resultado,

$$y(t) = x_1(t), \quad x_1(t) = \int_{t_0}^t (x_2(t) - \alpha_{m-1} y(\tau) + \beta_{m-1} u(\tau)) d\tau \quad (1.69)$$



**Figura 1.12.** Modelo de un sistema de dimensión múltiple.

El modelo del procesador analógico, que realiza estas transformaciones, se muestra en la Figura (1.12). La salida  $y(t)$ ,  $t \geq t_0$ , se determina si además de tener la entrada  $u(t)$ ,  $t \geq t_0$ , se fijan las condiciones iniciales  $x_k(t_0)$ ,  $k = 1, \dots, m$  de las salidas de los respectivos integradores. Por lo tanto, estas señales se pueden seleccionar en calidad de variables de estado del sistema dinámico descrito por la ecuación (1.38). En correspondencia con las expresiones (1.67), (1.68) y (1.69), las ecuaciones para estas variables de estado son:

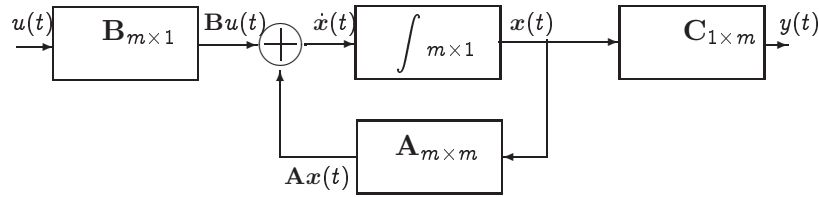
$$\begin{aligned} x_1(t) &= y(t) \\ x_2(t) &= \frac{dx_1}{dt} + \alpha_{m-1} y(t) - \beta_{m-1} u(t) \\ &\dots = \dots \\ x_m(t) &= \frac{dx_{m-1}}{dt} + \alpha_1 y(t) - \beta_1 u(t) \\ \frac{dx_m}{dt} &= -\alpha_0 y(t) + \beta_0 u(t) \end{aligned}$$

El sistema de ecuaciones de estado obtenidas para el sistema (1.38), al ser representadas mediante la ecuación equivalente de primer orden (1.61), implica de nuevo el cálculo de las correspondientes matrices en (1.63).

En particular, se obtiene que

$$\mathbf{A} = \begin{bmatrix} -\alpha_{m-1} & 1 & 0 & \cdots & 0 \\ -\alpha_{m-2} & 0 & 1 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ -\alpha_1 & 0 & 0 & \cdots & 1 \\ -\alpha_0 & 0 & 0 & \cdots & 0 \end{bmatrix}, \quad \mathbf{B} = \begin{bmatrix} \beta_{m-1} \\ \beta_{m-2} \\ \vdots \\ \beta_1 \\ \beta_0 \end{bmatrix}$$

La representación de la salida sigue siendo la misma que en (1.64). Aunque la ilustración del modelo de sistema de dimensión múltiple de la Figura 1.12, puede ser simplificada, empleando dispositivos matriciales con varias entradas y salidas, como se observa en la Figura 1.13.



**Figura 1.13.** Modelo matricial de un sistema lineal con dimensión  $m$

Las ecuaciones (1.62) y (1.64) se emplean para la representación de sistemas variantes, esto es, con parámetros variables en el tiempo, descritos en forma generalizada por la expresión (1.38). Sin embargo, en este caso las matrices (1.63) son dependientes del tiempo, por lo cual la descripción del modelo, asumiendo  $t \geq t_0$  y dado el vector de condiciones iniciales  $\mathbf{x}(t_0)$ , es:

$$\begin{cases} \frac{d\mathbf{x}}{dt} &= \mathbf{A}(t) \mathbf{x}(t) + \mathbf{B}(t) u(t) \\ y(t) &= \mathbf{C}(t) \mathbf{x}(t) \end{cases} \quad (1.70)$$

---

**Ejemplo 1.8.** Sea el par de sistemas unidos, como se muestra en la parte superior de la Figura 1.14(a), descritos por el modelo

$$\begin{cases} \frac{d\mathbf{x}_i}{dt} &= \mathbf{A}_i(t) \mathbf{x}_i(t) + \mathbf{B}_i(t) u_i(t) \\ y_i(t) &= \mathbf{C}_i(t) \mathbf{x}_i(t), \quad \forall t \geq t_0, \quad i \in \{1, 2\}, \quad \mathbf{x}_{i_{r_i \times 1}}(t), \quad r_i \in \{m, n\} \end{cases}$$

Hallar los parámetros del modelo conjunto (parte inferior de la Figura 1.14(a)).

La ecuación vectorial única del sistema conjunto tiene la misma estructura que (1.70) con los

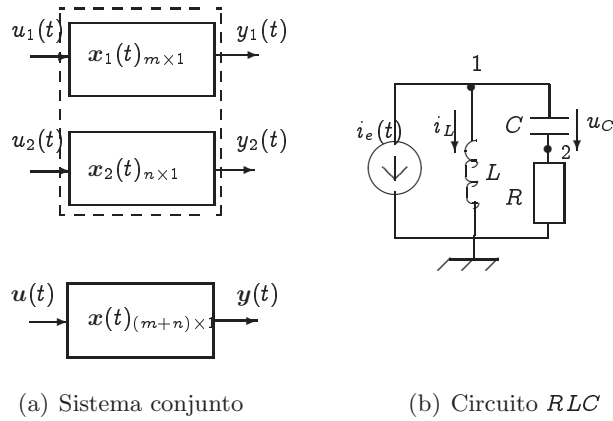


Figura 1.14. Ejemplo de sistemas

siguientes vectores únicos de estado y señales de entrada y salida, respectivamente:

$$\mathbf{x}(t) = \begin{bmatrix} \mathbf{x}_1(t) \\ \vdots \\ \mathbf{x}_2(t) \end{bmatrix}, \quad \mathbf{z}(t) = \begin{bmatrix} \mathbf{z}_1(t) \\ \vdots \\ \mathbf{z}_2(t) \end{bmatrix}, \quad \mathbf{z} \in \{\mathbf{u}, \mathbf{y}\}$$

mientras, la correspondiente estructura de matriz de parámetros tiene la forma,

$$\mathbf{P}(t) = \begin{bmatrix} \mathbf{P}_1(t) & \mathbf{0} \\ \mathbf{0} & \mathbf{P}_2(t) \end{bmatrix}, \quad \mathbf{P} \in \{\mathbf{A}, \mathbf{B}, \mathbf{C}\}, \quad \mathbf{P}_i \in \{\mathbf{A}_i, \mathbf{B}_i, \mathbf{C}_i\}, \quad i \in \{1, 2\}$$

**Ejemplo 1.9.** Hallar las correspondientes matrices del modelo de ecuación de estados del sistema para el circuito mostrado en la Figura 1.14(b).

La ecuación diferencial del circuito  $RLC$  1.14(b) corresponde a la expresión

$$LC \frac{d^2 i_L}{dt^2} + RC \frac{di_L}{dt} + i_L(t) = i_e(t) + RC \frac{di_e}{dt}$$

En calidad de variables de estado se escogen el voltaje en el condensador  $x_1 = u_C$  y la tensión en la bobina,  $x_2 = i_L$ , descritas respectivamente por las siguientes ecuaciones diferenciales:

$$\frac{du_C}{dt} = \frac{1}{C} (i_e(t) - i_L), \quad \frac{di_L}{dt} = \frac{1}{L} (u_C + Ri_e(t) - Ri_L)$$

De las anteriores ecuaciones, se observa que las matrices son constantes en el tiempo, motivo por el cual el modelo (1.70) se puede completar hasta la forma

$$\frac{d\mathbf{x}}{dt} = \mathbf{A}\mathbf{x}(t) + \mathbf{B}\mathbf{u}(t) \tag{1a}$$

$$\mathbf{y}(t) = \mathbf{C}\mathbf{x}(t) + \mathbf{D}\mathbf{u}(t) \tag{1b}$$

luego, para  $\mathbf{u}(t) = i_e(t)$ , se tienen las siguientes matrices:

$$\mathbf{A} = \begin{bmatrix} 0 & -1/C \\ 1/L & -R/L \end{bmatrix}, \quad \mathbf{B} = \begin{bmatrix} 1/C \\ R/L \end{bmatrix}$$



Si en calidad de variables de estado se escogen los voltajes  $y_i \in \{v_1, v_2\}$  sobre los puntos 1 y 2, mostrados en el circuito *RLC* de la Figura 1.14(b), entonces se tienen las ecuaciones:

$$\begin{cases} v_1 &= u_C + v_2 = u_C + R(i_e(t) - i_L), \\ v_2 &= R(i_e(t) - i_L) \end{cases}$$

con lo que las matrices respectivas se describen como  $\mathbf{C} = \begin{bmatrix} 1 & -R \\ 0 & -R \end{bmatrix}$ ,  $\mathbf{D} = \begin{bmatrix} R \\ R \end{bmatrix}$ .

### 1.4.2. Solución de la ecuación de estado

**Ecuación homogénea con coeficientes constantes.** Se asume la ecuación diferencial

$$\frac{d\mathbf{x}}{dt} = \mathbf{A}\mathbf{x}(t), t \geq t_0 \quad (1.72)$$

con condiciones iniciales  $\mathbf{x}(t_0)$ ,  $t \geq t_0$ . Cuando  $\mathbf{x}(t) \Rightarrow \mathbf{x}(t)$ , esto es, se considera una función escalar, para (1.72) se tiene la solución,  $\mathbf{x}(t) = e^{\mathbf{A}(t-t_0)}\mathbf{x}(t_0)$ . La misma estructura de solución se considera para el caso,  $\mathbf{x}(t) = e^{\mathbf{A}(t-t_0)}\mathbf{x}(t_0)$ , considerando la expansión:  $e^{\mathbf{A}t} = \mathbf{I} + \mathbf{A}t + \frac{(\mathbf{A}t)^2}{2!} + \dots + \frac{(\mathbf{A}t)^n}{n!}$ . Además, si la matriz  $\mathbf{A}$  es de orden  $n \times n$ , entonces cualquier polinomio o serie convergente de  $\mathbf{A}$ , puede ser descrita en forma de combinación lineal de los términos  $\mathbf{I}, \mathbf{A}, \dots, \mathbf{A}^n$ .

Sea la notación de la función matricial  $\Phi(t-t_0) \triangleq e^{\mathbf{A}(t-t_0)}$ , denominada *matriz de transición de estado*, y de la cual se exigen las siguientes condiciones:

$$\begin{cases} \Phi(t_0 - t_0) &= \mathbf{I} \\ \frac{d}{dt}\Phi(t-t_0) &= \mathbf{A}(t)\Phi(t-t_0) \\ \frac{d}{dt}\Phi^\top(t, t_0) &= -\mathbf{A}^\top(t_0)\Phi(t, t_0) \end{cases}$$

Por lo tanto, la solución del caso homogéneo tiene forma  $\mathbf{x}(t) = \Phi(t-t_0)\mathbf{x}(t_0)$ .

**Ecuación homogénea con coeficientes variables.** Entonces, las matrices del modelo (1.70) son variantes en el tiempo. Se considera la matriz de transición de estado como función de dos variables,  $\Phi(t, t_0) = e^{\mathbf{A}(t, t_0)}$ , que cumple la siguiente ecuación diferencial:

$$\frac{d}{dt}\Phi(t, t_0) = \mathbf{A}(t)\Phi(t, t_0), \quad \Phi(t_0, t_0) = \mathbf{I}$$

con las siguientes propiedades adicionales, para  $t_0 < t_1 < t_2$ ,

$$\begin{cases} \Phi(t_2, t_0) &= \Phi(t_2, t_1)\Phi(t_1, t_0), \\ \Phi^{-1}(t_1, t_0) &= \Phi(t_0, t_1) \end{cases}$$

En realidad, una expresión compacta para  $\Phi(t, t_0)$  es difícil de obtener. Sin embargo, en la práctica, es más importante saber que ésta expresión existe y que posee ciertas propiedades.

Si es necesario hallar la matriz de transición de estado, se recurre a métodos numéricos. En general, la solución del caso homogéneo con parámetros variables se representa mediante la expresión,  $\mathbf{x}(t) = \Phi(t, t_0)\mathbf{x}(t_0)$ ,  $t \geq t_0$ .

**Ecuación no homogénea con coeficientes variables.** La solución de (1.62) es:

$$\mathbf{x}(t) = \Phi(t, t_0)\mathbf{x}(t_0) + \int_{t_0}^t \Phi(t, \tau)\mathbf{B}(\tau)\mathbf{u}(\tau)d\tau \quad (1.73)$$

donde la matriz de transición de estado  $\Phi(t, t_0)$  cumple las condiciones [11]:

$$\begin{aligned} \Phi(t_0, t_0) &= \mathbf{I}, (\Phi(t, \tau))^{-1} = \Phi(\tau, t), \\ \frac{d}{dt}\Phi(t, \tau) &= \mathbf{A}(t)\Phi(t, \tau) \\ \frac{d}{dt}\Phi^\top(t, \tau) &= -\mathbf{A}^\top(\tau)\Phi^\top(t, \tau) \\ \Phi(t, \tau)\Phi(\tau, \theta) &= \Phi(t, \theta) \end{aligned}$$

La solución (1.73) contiene dos componentes: la primera, que corresponde a las oscilaciones libres, y la segunda, que corresponde a las oscilaciones forzadas. Por cierto, para que el sistema se mantenga estable, el primer término debe atenuarse en el tiempo, de tal manera que el vector de estado se determine, más bien, por la componente forzada, que se obtiene como solución del mismo modelo (1.62), pero con condiciones iniciales 0.

El cálculo de la matriz  $\Phi(t)$  se realiza empleando la transformada de Laplace, (1.25a), sobre (1a), con lo que se obtiene la siguiente ecuación algebraica vectorial,

$$s\mathbf{X}(s) - \mathbf{x}(t_0) = \mathbf{A}^{-1}\mathbf{X}(s) + \mathbf{B}\mathbf{U}(s)$$

Entonces, la imagen del vector de las variables de estado es

$$\mathbf{X}(s) - \mathbf{x}(t_0) = (s\mathbf{I} - \mathbf{A})^{-1}\mathbf{x}(t_0) + (s\mathbf{I} - \mathbf{A})^{-1}\mathbf{B}\mathbf{U}(s)$$

luego,  $\mathbf{x}(t) = \mathcal{L}^{-1}\{(s\mathbf{I} - \mathbf{A})^{-1}\mathbf{x}(t_0)\} + \mathcal{L}^{-1}\{(s\mathbf{I} - \mathbf{A})^{-1}\mathbf{B}\mathbf{U}(s)\}$ .

Asumiendo la invariabilidad de las matrices, entonces la matriz de transición de estado se calcula como,  $\Phi(t) = \mathcal{L}^{-1}\{(s\mathbf{I} - \mathbf{A})^{-1}\}$ .

**Ejemplo 1.10.** Calcular la matriz de transición de estado para un sistema de segundo orden.

Los sistemas lineales se caracterizan por los valores propios  $\lambda_k, k = 1, \dots, n$  de la matriz  $\mathbf{A}$ , que corresponden a las raíces del polinomio característico,  $\det(\lambda\mathbf{I} - \mathbf{A}) = 0$ . Cuando todas las raíces del polinomio son diferentes,  $\lambda_k \neq \lambda_l, \forall k, l = 1, \dots, n$  entonces la matriz de transición de estado es igual a

$$\Phi(t) = e^{\lambda_1 t} \frac{\mathbf{A} - \lambda_2 \mathbf{I}}{\lambda_1 - \lambda_2} - e^{\lambda_2 t} \frac{\mathbf{A} - \lambda_1 \mathbf{I}}{\lambda_1 - \lambda_2}$$

En particular, sea el sistema de segundo orden,  $n = 2$ , descrito por el par de ecuaciones diferenciales  $dx_1/dt = x_2$ ,  $dx_2/dt = -a_1x_1 - a_2x_2 + u(t)$ , con lo que se obtienen las matrices

$$\mathbf{A} = \begin{bmatrix} 0 & 1 \\ -a_1 & -a_2 \end{bmatrix}, \mathbf{B} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

Los valores propios del sistema se determinan de la ecuación,  $\det(\lambda\mathbf{I} - \mathbf{A}) = \lambda^2 + a_2\lambda + a_1 = 0$ , que tiene la solución  $\lambda_{1,2} = -a_2/2 \pm \sqrt{a_2^2/4 - a_1}$ .

El reemplazo de los valores obtenidos para los valores propios en la matriz de transición de estados, en el caso particular de un sistema de segundo orden, da como resultado,

$$\Phi(t) = e^{-a_2t/2} \begin{bmatrix} \frac{a_2}{2\beta} \sin\beta t + \cos\beta t & \frac{1}{\beta} \sin\beta t \\ -\frac{a_1}{\beta} \sin\beta t & -\frac{a_2}{2\beta} \sin\beta t + \cos\beta t \end{bmatrix}, \quad \beta = \sqrt{a_1 - a_2^2/4}$$

Los sistemas lineales variantes en el tiempo con múltiples entradas y salidas, a partir de (1.35), se describen por la matriz cuadrada de la respuesta impulso  $\mathbf{h}(t, \tau)$  en la forma,

$$\mathbf{y}(t) = \int_{-\infty}^t \mathbf{h}(t, \tau) \mathbf{u}(\tau) d\tau \quad (1.74)$$

En muchos casos reales, cuando  $t_0 \rightarrow -\infty$ , en (1.73) se puede despreciar el primer término, con lo cual de (1.62) y (1.73) se obtiene,

$$\mathbf{y}(t) = \mathbf{C}(t) \int_{-\infty}^t \Phi(t, \tau) \mathbf{B}(\tau) \mathbf{u}(\tau) d\tau \quad (1.75)$$

Así mismo, de las expresiones (1.74) y (1.75) se deduce que,

$$\mathbf{h}(t, \tau) = \begin{cases} \mathbf{C}(t) \Phi(t, \tau) \mathbf{B}(\tau), & \tau \leq t \\ \mathbf{0}, & \tau > t \end{cases} \quad (1.76)$$

Cabe anotar que las matrices, que determinan la respuesta a impulso en (1.76), dependen del método elegido de disposición del vector de estado del sistema considerado en cada caso concreto, y por lo tanto, para un mismo sistema las matrices pueden ser diferentes. Sin embargo, la respuesta a impulso matricial  $\mathbf{h}(t, \tau)$  es única para un sistema dado.

En general, existe una transformación lineal de las coordenadas de estado, en la forma  $\mathbf{z} = \mathbf{F}\mathbf{x}$ , tal que del sistema conformado por las ecuaciones (1.62) y (1.64) se obtiene el modelo,

$$\begin{cases} \dot{\mathbf{z}} = \hat{\mathbf{A}}\mathbf{z} + \mathbf{B}\mathbf{u} \\ \mathbf{y} = \hat{\mathbf{C}}^T \mathbf{z} \end{cases}$$

donde

$$\hat{\mathbf{A}} = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & \cdots & 1 \\ -a_n & -a_{n-1} & -a_{n-2} & \cdots & -a_1 \end{bmatrix}, \hat{\mathbf{B}} = \mathbf{F}\mathbf{B}, \hat{\mathbf{C}} = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \mathbf{F} = \begin{bmatrix} \mathbf{C}^\top \\ \mathbf{C}^\top \mathbf{A} \\ \vdots \\ \mathbf{C}^\top \mathbf{A}^{n-1} \end{bmatrix} \quad (1.77)$$

En concordancia con los valores de las matrices en (1.77), el modelo (1.62) se puede reemplazar por la ecuación diferencial escalar del tipo,

$$\frac{d^n y}{dt^n} + a_1 \frac{d^{n-1} y}{dt^{n-1}} + \cdots + a_n y = b_1 \frac{d^{n-1} u}{dt^{n-1}} + \cdots + b_n u, \quad b_i = \sum_{k=0}^{i-1} a_{i-k} \hat{g}_k + \hat{g}_i \quad (1.78)$$

En los sistemas lineales, cuando las matrices  $\mathbf{A}(t)$ ,  $\mathbf{B}(t)$  y  $\mathbf{C}(t)$  muestran dependencia del tiempo, se puede encontrar la correspondiente representación en forma de ecuación diferencial (1.78), con los siguientes coeficientes:

$$b_i(t) = \sum_{k=0}^{i-1} \sum_{l=0}^{i-k} C_{n+l-1}^{n-1} a_{i-k-l} \frac{d}{dt^l} \hat{g}_k(t) + \hat{g}_i(t)$$

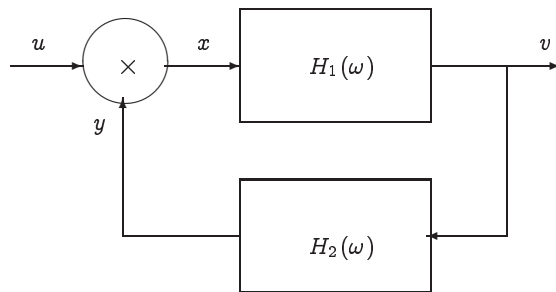
La representación del sistema por las expresiones (1.77) y (1.78), frecuentemente es más cómoda para su análisis mediante procesadores digitales.

## Problemas

**Problema 1.20.** Hallar el modelo equivalente de los circuitos mostrados en la Figura 1.8.

**Problema 1.21.** Hallar la matriz de transición de estados de los circuitos de la Figura 1.8.

**Problema 1.22.** Hallar la estabilidad del sistema con retroalimentación mostrado en la Figura 1.15, con las siguiente funciones de transferencia:  $H_1(\omega) = k/(1 + j\omega)^3$  y  $H_2(\omega) = 1$ .



**Figura 1.15.** Ejemplo de sistema con retroalimentación

# Capítulo 2

## Representación de variables aleatorias

El concepto general de medida  
comprende, como caso particular, el  
concepto de probabilidad

---

Kolmogorov

El desarrollo de la teoría de funciones en espacios métricos implica el estudio de aquellas propiedades de las funciones, las cuales dependen de la medida de conjuntos en los que las funciones toman cualquier grupo de valores. Así por ejemplo, se tienen las propiedades de ortogonalidad de un conjunto de funciones y su completitud, analizadas en §1.1.2. Si se consideran los valores aleatorios como funciones, determinadas en un espacio de eventos elementales, entonces todos métodos de Teoría de Probabilidades aplicadas a estas funciones son un caso específico del empleo de la Teoría Generalizada de Medida. En particular, gran parte de los métodos de análisis probabilístico se basan en el empleo del concepto de independencia de variables aleatorias, que se puede entender como una propiedad de los espacios métricos.

### 2.1. Valores y funciones de probabilidad

La aleatoriedad implica incertidumbre. La representación de las variables aleatorias consiste en la descripción y caracterización de sus principales propiedades que midan esa incertidumbre inherente.

#### 2.1.1. Espacio probabilístico de variables

Cualquier modelo de un experimento analizado a priori se caracteriza por un conjunto de posibles resultados u *observaciones*. El subconjunto formado por un solo elemento se denomina *evento* o *suceso elemental* y todo el conjunto  $\Omega$  se denomina *espacio de eventos elementales*. Los posibles subconjuntos de eventos  $A_k \in \Omega$ , pueden ser de dos clases: eventos

incompatibles, cuando:

$$\bigcup_{i=1}^n A_i = \Omega, \quad A_k \cap A_l = \emptyset, \quad k, l = 1, \dots, n$$

y eventos complementarios,  $\{\bar{A}_k\}$ , tales que

$$A \cup \bar{A} = \Omega \text{ y } A \cap \bar{A} = \emptyset$$

**Definición 2.1.** ( $\sigma$ -álgebra  $\mathcal{B}$ ) El sistema conformado por los subconjuntos de  $\Omega$  se denomina álgebra  $\mathcal{B}$ , si se cumple que para  $\{A_i \in \mathcal{B} : i = 1, \dots, n\}$  se observa la igualdad

$$\bigcup_{i=1}^n A_i = \mathcal{B}$$

esto es, de la sucesión de eventos  $A_i$ , por lo menos ocurre una. Así mismo, se cumple que

$$\bigcap_{i=1}^n A_i = \mathcal{B}$$

por lo que todos los sucesos  $A_k$  ocurren de manera simultánea. La definición se completa con la condición por la cual teniendo  $A \in \mathcal{B}$ , resulta que  $\bar{A} \in \mathcal{B}$ . Por lo anterior, el álgebra  $\mathcal{B}$  es una clase de conjuntos cerrados con relación a una cantidad contable de las operaciones básicas de unión, intersección y complemento ( $\sigma$ -álgebra).

En general, el *espacio probabilístico*  $\{\Omega, \mathcal{B}, P\}$  está conformado por las siguientes tres componentes: el espacio de los eventos elementales  $\Omega$ , la  $\sigma$ -álgebra  $\mathcal{B}$  de los subconjuntos de Borell del espacio  $\Omega$  ( $\sigma$ -álgebra de los eventos) y la estabilidad de las frecuencias de aparición en términos relativos de los eventos o medida de *probabilidad*  $P(A)$ , determinada para  $A \in \mathcal{B}$ , con una masa total unitaria, esto es,  $P(\Omega) = 1$ . La medida de probabilidad se determina a partir de los siguientes principios, denominados *axiomas de Kolmogorov* [12]:

**Axioma I.** A cada evento  $A \in \mathcal{B}$  le corresponde un valor real  $P(A)$ , denominado probabilidad de  $A$ , tal que,  $0 \leq P(A) \leq 1$ .

**Axioma II.**  $P(\Omega) = 1$ .

**Axioma III.** Si  $A \cap B = \emptyset$ , entonces,  $P(A \cup B) = P(A) + P(B)$ .

De los anteriores axiomas se deducen las siguientes propiedades:

- (a). Si  $A$  es un evento imposible, entonces  $P(A) = 0$ .
- (b).  $P(\bar{A}) = 1 - P(A)$ .
- (c). Si  $A \subseteq B$  entonces, se cumple la desigualdad,  $P(A) \leq P(B)$ .
- (d). Si se cumple que  $A_i \cap A_k = \emptyset$ ;  $i, k = 1, \dots, n$ , luego,

$$P\left(\bigcup_{k=1}^n A_k\right) = \sum_{k=1}^n P(A_k) \tag{2.1}$$

En forma general, si  $A_i \cap A_k \neq \emptyset$ ;  $i, k = 1, \dots, n$ , entonces,

$$P\left(\bigcup_{k=1}^n A_k\right) \leq \sum_{k=1}^n P(A_k)$$

(e). Dada cualquier pareja de sucesos  $A$  y  $B$ , se cumple

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

Dos eventos  $A$  y  $B$  son *dependientes* si la probabilidad  $P(A)$  depende de que suceda el evento  $B$ . La probabilidad de la aparición conjunta de ambos eventos se da como

$$P(A \cap B) = P(A)P(B|A) = P(B)P(A|B)$$

Las probabilidades marginales  $P(A)$  y  $P(B)$  se denominan *a priori*, mientras las condicionales  $P(B|A)$  y  $P(A|B)$  - *a posteriori*.

Cuando se asume la independencia mutua de aparición de cada uno de los eventos  $A$  y  $B$ , las respectivas probabilidades a priori y a posteriori se determinan por las expresiones:  $P(B|A) = P(B)$  y  $P(A|B) = P(A)$ , y por lo tanto,  $P(A \cap B) = P(A)P(B)$ , relación que se puede generalizar en la definición de *independencia* del conjunto de sucesos  $\{A_k\}$ :

$$P\left(\bigcap_{k=1}^n A_k\right) = \prod_{k=1}^n P(A_k) \quad (2.2)$$

En cambio, si el conjunto  $\{A_k\}$  es *dependiente* se cumple que:

$$P\left(\bigcap_{k=1}^n A_k\right) = P(A_1)P(A_2|A_1)P(A_3|A_1 \cap A_2) \cdots P\left(A_n \middle| \bigcap_{k=1}^{n-1} A_k\right)$$

**Ejemplo 2.1.** Se reciben 10 números en una serie aleatoria con valores equiprobables del 0 al 9, generados de forma independiente. Hallar la probabilidad de que el primer número obtenido de la serie sea múltiplo de 3.

Los posibles eventos  $A_k$  son 3 (3, 6 y 9), ambos dados con probabilidad de generación igual a  $P(A_k) = 0.1$ , son independientes y excluyentes, por lo tanto, la aparición de cada uno de ellos, acordes con la expresión (2.1), es igual a

$$P(3, 6, 9) = 0.1 + 0.1 + 0.1 = 0.3.$$

**Ejemplo 2.2.** Del anterior ejemplo, hallar la probabilidad de aparición de que por lo menos uno de los valores de la serie aleatoria sea múltiplo de 3.

En este caso, los eventos son independientes pero no excluyentes. La aparición de cualquier suceso  $A_k$  excluye la aparición de la posibilidad de no aparición de todos los demás eventos; situación que se puede analizar como el complemento de (2.2),

$$P\left(\bigcup_{k=1}^n A_k\right) = 1 - P\left(\bigcap_{k=1}^n \bar{A}_k\right) = \prod_{k=1}^n P(\bar{A}_k) = 1 - \prod_{k=1}^n (1 - P(A_k))$$

que en el caso particular de  $n = 10$  y  $P(A_k) = 1/n$ , resulta en  $P(\cup_{k=1}^n A_k) = 0.271$ .

La medida de probabilidad  $P(A)$  se completa con los siguientes dos teoremas:

**Teorema 2.2.** (Probabilidad completa) La probabilidad del evento  $A$ , de que aparezca con una de las  $n$  posibles hipótesis  $B_1, \dots, B_n$  o sucesos incompatibles que conforman un grupo topológico completo, esto es,  $A = \bigcup_{k=1}^n (A \cap B_k)$  y teniendo en cuenta que  $(A \cap B_k) \cap (A \cap B_l) = \emptyset, \forall k \neq l$ , se define como

$$\begin{aligned} P(A) &= \sum_{k=1}^n P\left(\bigcup_{k=1}^n (A \cap B_k)\right) = \sum_{k=1}^n P(A \cap B_k) \\ &= \sum_{k=1}^n P(B_k) P(A|B_k) \end{aligned} \quad (2.3)$$

**Teorema 2.3.** (Bayes) Sea el conjunto de hipótesis  $\{B_k : k = 1, \dots, n\}$  un grupo completo, empleando la regla de multiplicación se encuentra que

$$P(B_k|A) = \frac{P(B_k) P(A|B_k)}{P(A)}$$

Si se reemplaza la probabilidad a priori  $P(A)$  por su valor dado en (2.3), se obtiene la siguiente expresión, denominada fórmula de Bayes:

$$P(B_k|A) = \frac{P(B_k) P(A|B_k)}{\sum_{k=1}^n P(B_k) P(A|B_k)} \quad (2.4)$$

**Ejemplo 2.3.** Las resistencias de un sistema se obtienen de dos puntos de producción diferentes: la fábrica  $A_1$  (que produce el 70 %) y la fábrica  $A_2$  (30 %). Se conoce que la fábrica  $A_1$  en promedio produce el 83 % de resistencias con calidad aceptable, mientras la fábrica  $A_2$ , tan solo el 63 %. Determinar las siguientes probabilidades:

1. La probabilidad de que la primera resistencia escogida al azar sea de cada una de las fábricas:  $A_1$  y  $A_2$ .
2. La probabilidad de que las dos primeras resistencias escogidas al azar sean de cada una de las fábricas:  $A_1$  y  $A_2$ .
3. La probabilidad de que por lo menos una resistencia de las dos primeras escogidas al azar sean de cada una de las fábricas:  $A_1$  y  $A_2$ .
4. La probabilidad de que la primera resistencia escogida al azar sea de calidad y producida por cada una de las fábricas:  $A_1$  y  $A_2$ .
5. La probabilidad de que la primera resistencia escogida al azar tenga calidad aceptable.
6. La probabilidad de que si la primera resistencia escogida al azar tiene la calidad aceptable, entonces sea de cada una de las fábricas:  $A_1$  y  $A_2$ .

Los valores respectivos de probabilidad se hallan para cada caso de la siguiente forma,

**caso 1.** Se tiene un solo evento, que por definición es igual a

$$P(A_1) = \frac{70}{100} = 0.7, \quad P(A_2) = 0.3$$

**caso 2.** Se tienen dos eventos independientes y no excluyentes. La probabilidad de que ambos ocurran se calcula por la expresión (2.2),

$$\begin{aligned} P(A_1(1) \cap A_1(2)) &= 0.7 \cdot 0.7 = 0.49 \\ P(A_2(1) \cap A_2(2)) &= 0.3 \cdot 0.3 = 0.09 \end{aligned}$$



**caso 3.** La probabilidad de que se tenga por lo menos una resistencia de dos hechas por una de las fábricas,  $A_1$  y  $A_2$ , se calcula por la expresión obtenida en el ejemplo 2.2, respectivamente como:

$$P(A_1(1) \cup A_1(2)) = 1 - (1 - P(A_1))(1 - P(A_1)) = 1 - (1 - 0.7)^2 = 0.91$$

$$P(A_2(1) \cup A_2(2)) = 1 - (1 - 0.3)^2 = 0.09$$

**caso 4.** Los eventos considerados son dependientes y no excluyentes; la resistencia de buena calidad  $c$  hecha por cada una de las fábricas  $A_i$ ,  $i = 1, 2$ , con probabilidad condicional  $P_{A_i}(c)$ , cuyo valor de probabilidad se halla respectivamente por la expresión:

$$P(A_1 \cup c) = P(A_1) P_{A_1}(c) = 0.7 \cdot 0.83 = 0.581$$

De igual manera se tiene,  $P(A_2 \cup c) = P(A_2) P_{A_2}(c) = 0.3 \cdot 0.63 = 0.189$ .

**caso 5.** En el caso de selección de una resistencia con aceptable calidad  $c$ , como hipótesis  $B_k$  se asume que la resistencia se hizo en cada una de las fábricas:  $A_i$  ( $P(B_k) = P(A_k)$ , para  $k = 1, 2$ ). Empleando la expresión de probabilidad completa (2.3) se tiene,

$$\begin{aligned} P(c) &= P(A_1 \cap c \cup A_2 \cap c) = P(B_1) P_{A_1}(c) + P(B_2) P_{A_2}(c) \\ &= 0.7 \cdot 0.83 + 0.3 \cdot 0.63 = 0.77 \end{aligned}$$

**caso 6.** El valor de la probabilidad de la hipótesis  $B_k$  (la resistencia se hizo en la fábrica  $A_k$ ) con la condición de que la resistencia se termina con calidad aceptable se calcula por el teorema de Bayes (2.4),

$$\begin{aligned} P_c(A_1) &= \frac{P(A_1) P_{A_1}(c)}{P(A_1) P_{A_1}(c) + P(A_2) P_{A_2}(c)} = \frac{0.7 \cdot 0.83}{0.7 \cdot 0.83 + 0.3 \cdot 0.63} = 0.755 \\ P_c(A_2) &= \frac{P(A_2) P_{A_2}(c)}{P(A_1) P_{A_1}(c) + P(A_2) P_{A_2}(c)} = \frac{0.3 \cdot 0.63}{0.7 \cdot 0.83 + 0.3 \cdot 0.63} = 0.245 \end{aligned}$$

Cuando se establece alguna regla o relación funcional que asigna un número real a cada posible observación o trayectoria de la *variable aleatoria asignada*, la totalidad de sus valores posibles constituye el conjunto  $\{x_i \in \mathbb{R} : i = 1, 2, \dots\}$  confinado en un espacio  $\mathfrak{X} \in \mathbb{R}$  y conocido como espacio de observaciones o *muestra*. Las variables aleatorias pueden ser discretas o continuas dependiendo de la naturaleza de las observaciones del experimento. Si en cualquier intervalo finito del eje real, la variable  $\xi$  puede asumir solamente un número finito de valores distintos y a priori conocidos, se conoce a la variable como *aleatoria discreta*. Si, por el contrario,  $\xi$  puede tomar cualquier valor en un intervalo dado del eje real, se denomina *variable aleatoria continua*. En este caso, en cualquier intervalo del eje real la cantidad de valores que puede tomar  $\xi$  es infinita.

Un valor aleatorio en  $\{\Omega, \mathcal{B}, P\}$  corresponde a cualquier  $\mathcal{B}$ -función medible  $\xi(\omega) \in \Omega$  (se nota simplemente por  $\xi$ ). La distribución de la variable aleatoria  $\xi$  corresponde a la medida  $F_\xi(x) = P_\xi(\xi < x : \xi \in \mathcal{B})$ . Cuando se cumple que el valor de probabilidad

$$P(\xi = x) = F_\xi(x + \Delta x) - F_\xi(x) > 0, \quad \Delta x \rightarrow 0,$$

entonces  $x$  se denomina el *átomo* de la distribución.

La variable  $\xi$  tiene distribución continua y absoluta si existe la función  $p_\xi > 0$ , que

sea normalizada en el sentido en que  $\int_{\Lambda} p_{\xi}(\lambda) d\lambda = 1$ , denominada *función densidad de probabilidad* (FDP), para la cual se cumple que:

$$\begin{cases} F_{\xi}(x) &= \int_{-\infty}^x p_{\xi}(\lambda) d\lambda, \\ p_{\xi}(\lambda) &= dF_{\xi}(\lambda)/d\lambda \end{cases} \quad (2.5)$$

Se considera que una variable aleatoria  $\xi_n$ ,  $n = 1, 2, \dots$ , converge en el sentido probabilístico a un valor o función  $\xi$ , si para cualquier valor tan pequeño como sea de  $\varepsilon > 0$ , la probabilidad de ocurrencia de la desigualdad corresponde al límite

$$\lim_{\substack{n \rightarrow \infty \\ \varepsilon \rightarrow 0}} P(|\xi_n - \xi| < \varepsilon) = 1, \quad \forall \varepsilon > 0$$

### 2.1.2. Valores medios y momentos

Una forma adecuada para describir una variable aleatoria  $\xi$  corresponde a los valores y momentos que determinan las particularidades de su respectiva FDP,  $p_{\xi}(x)$ .

Los *momentos iniciales* se definen por la siguiente expresión:

$$E\{\xi^n\} \triangleq \int_{-\infty}^{\infty} \xi^n dF_{\xi}(x) = \int_{-\infty}^{\infty} x^n p_{\xi}(x) dx = m_{n\xi}, \quad n \in \mathbb{N} \quad (2.6)$$

**Momento de primer orden.** Valor medio conocido como la *esperanza matemática*:

$$m_{1\xi} = E\{\xi\} = \int_{-\infty}^{\infty} x p_{\xi}(x) dx \quad (2.7)$$

Si la variable aleatoria es discreta, la integral (2.7) se reduce a la sumatoria en la forma:  $m_{1\xi} = \sum_i^N x_i P(\xi = x_i)$ . Por cierto, cuando los eventos son equiprobables,  $x_i = 1/N, \forall i$ , entonces, el primer momento inicial corresponde al *promedio*.

Las siguientes son las principales propiedades del valor medio

$$\begin{cases} E\{c\} &= c, \quad c = \text{const} \\ E\{c\xi\} &= cm_{1\xi}, \quad c = \text{const} \\ E\{\xi \pm \eta\} &= m_{1\xi} \pm m_{1\eta} \end{cases}$$

El concepto de esperanza matemática de la variable aleatoria  $\xi$  se puede generalizar, para el caso de cualquier función determinística  $f(\xi)$ , en la forma:

$$E\{f(\xi)\} = \int_{-\infty}^{\infty} f(x) dF_{\xi}(x) = \int_{-\infty}^{\infty} f(x) p_{\xi}(x) dx \quad (2.8)$$

De otra parte, a fin de evitar la influencia que tiene el valor medio en el cálculo de los momentos de orden superior  $n > 1$  se definen los *momentos centralizados* de orden  $n$ :

$$E\{(\xi - m_{1\xi})^n\} = \int_{-\infty}^{\infty} (x - m_{1\xi})^n p_{\xi}(x) dx = \mu_{n\xi}, \quad \forall n \geq 2, n \in \mathbb{N} \quad (2.9)$$

**Momentos de segundo orden.** Corresponde al *valor cuadrático medio*,

$$m_{2\xi} = \int_{-\infty}^{\infty} x^2 p_{\xi}(x) dx, \quad (2.10)$$

De forma análoga, la sumatoria  $m_{2\xi} = \sum_i^N x_i^2 P_{\xi}(\xi = x_i)$  determina este momento para las variables discretas.

El momento central (2.9) para  $n = 2$  se conoce como *varianza*:

$$\begin{aligned} \mu_{2\xi} &= \int_{-\infty}^{\infty} (x - m_{1\xi})^2 p_{\xi}(x) dx = \int_{-\infty}^{\infty} (x^2 - 2xm_{1\xi} + m_{1\xi}^2) p_{\xi}(x) dx \\ &= \int_{-\infty}^{\infty} (x^2) p_{\xi}(x) dx - 2m_{1\xi} \int_{-\infty}^{\infty} (x) p_{\xi}(x) dx + m_{1\xi}^2 \int_{-\infty}^{\infty} p_{\xi}(x) dx \end{aligned}$$

Teniendo en cuenta las definiciones, (2.7) y (2.10), se obtiene

$$\mu_{2\xi} = m_{2\xi} - 2m_{1\xi}^2 + m_{1\xi}^2 = m_{2\xi} - m_{1\xi}^2 \triangleq \sigma_{\xi}^2 \quad (2.11)$$

Si la magnitud aleatoria es discreta, cuando ésta toma los valores  $\xi_k$  con las probabilidades asociadas  $P(\xi_k)$ , a partir de (2.11), la respectiva varianza se calcula como:

$$\sigma_{\xi}^2 = \sum_{k=1}^N x_k^2 P_{\xi}(\xi = x_k) - \left( \sum_{k=1}^N x_k P_{\xi}(\xi = x_k) \right)^2$$

El valor  $\sigma_{\xi}^2$  caracteriza el nivel de concentración de la FDP,  $p_{\xi}(x)$ , en los alrededores de la esperanza matemática; interpretación que resulta de la *desigualdad de Chebyshev*:

$$P_{\xi}(|x - m_{1\xi}| \geq \varepsilon) \leq \sigma_{\xi}^2 / \varepsilon^2 \quad (2.12)$$

esto es, al disminuir el nivel de dispersión de la variable aleatoria  $\xi$ , aumenta la probabilidad de que sus valores no salgan fuera de los límites del intervalo  $[m_{1\xi} - \varepsilon, m_{1\xi} + \varepsilon]$ , siendo

$\varepsilon \geq 0$ . Cabe anotar que si la FDP es simétrica,  $p_\xi(x) = p_\xi(-x)$ , entonces,  $E\{\xi\} = 0$ , luego,

$$\begin{aligned}\sigma_\xi^2 &= \int_{-\infty}^{\infty} (x - m_{1\xi})^2 p_\xi(x) dx = \int_{-\infty}^{\infty} x^2 p_\xi(x) dx \\ &= 2 \int_0^{\infty} x^2 p_\xi(x) dx\end{aligned}\quad (2.13)$$

En general, la varianza tiene las siguientes propiedades:

$$\begin{cases} \sigma_\xi^2 \geq 0, \text{ si } \sigma_\xi^2 = 0 \Rightarrow \xi \rightarrow \text{const.} \\ \sigma_{c\xi}^2 = \int_{-\infty}^{\infty} (cx - cm_{1\xi})^2 p_\xi(x) dx = c^2 \sigma_\xi^2, \text{ por cierto, } \sigma_{\xi+c}^2 = \sigma_\xi^2, \forall c = \text{const.} \\ \sigma_{\xi\eta}^2 = E\{\xi^2 \eta^2\} = E\{\xi^2\} E\{\eta^2\} = \sigma_\xi^2 \sigma_\eta^2 \end{cases} \quad (2.14)$$

**Ejemplo 2.4.** Una moneda se lanza tres veces. Determinar la función acumulativa correspondiente a todos los posibles sucesos.

De forma condicional, cuando la moneda muestre la cara se denota por 0 y cuando muestre el sello, por 1. Asegurando la independencia de los lanzamientos, los sucesos se agrupan de acuerdo a la cantidad de caras y sellos que muestren sin importar su orden de salida: suceso cero: que haya 3 caras y ningún sello; ocurriría una vez de ocho posibles (1/8); suceso uno: que haya 2 caras y un sello; ocurriría tres veces de ocho posibles (3/8); suceso dos: que haya 1 cara y 2 sellos; saldría tres veces de ocho posibles (3/8); suceso tres: que no haya ninguna cara y tres sellos: una vez de ocho posibles (1/8). De acuerdo al orden descrito de los sucesos, la variable aleatoria toma los valores 0, 1, 2, 3, con función distributiva (representada en la parte inferior de la Figura 2.1(a)) descrita como:

$$F_\xi(x) = (u(x) + 3u(x-1) + 3u(x-2) + u(x-3))/8$$

La FDP respectiva está representada en la parte superior de la Figura 2.1(a) que corresponde a la derivada de la función distributiva e igual a:

$$p_\xi(x) = (\delta(x) + 3\delta(x-1) + 3\delta(x-2) + \delta(x-3))/8$$

**Ejemplo 2.5.** Dada la señal binaria periódica aleatoria, representada en la parte superior de la Figura 2.1(b). Hallar los 2 primeros momentos iniciales y la varianza.

Por cuanto,  $p_\xi(x) = P_1\delta(x-x_1) + P_2\delta(x-x_2)$ , donde  $P_1 = 1 - P_2$ , entonces, los respectivos momentos se calculan como:

$$\begin{aligned}m_{1\xi} &= \int_{-\infty}^{\infty} x (P_1\delta(x-x_1) + P_2\delta(x-x_2)) dx = P_1x_1 + P_2x_2; \\ m_{2\xi} &= \int_{-\infty}^{\infty} x^2 p_\xi(x) dx = \int_{-\infty}^{\infty} x^2 (P_1\delta(x-x_1) + P_2\delta(x-x_2)) dx = P_1x_1^2 + P_2x_2^2; \\ \sigma_\xi^2 &= P_1P_2(x_1-x_2)^2 = P_1(1-P_1)(x_1-x_2)^2\end{aligned}$$

Cuando  $P_1 = P_2 = 0.5 \Rightarrow \sigma_\xi^2 = (x_1 - x_2)^2/4$ , asumiendo  $\xi_1 = -\xi_2 = 1 \Rightarrow \sigma^2 = 1$ .

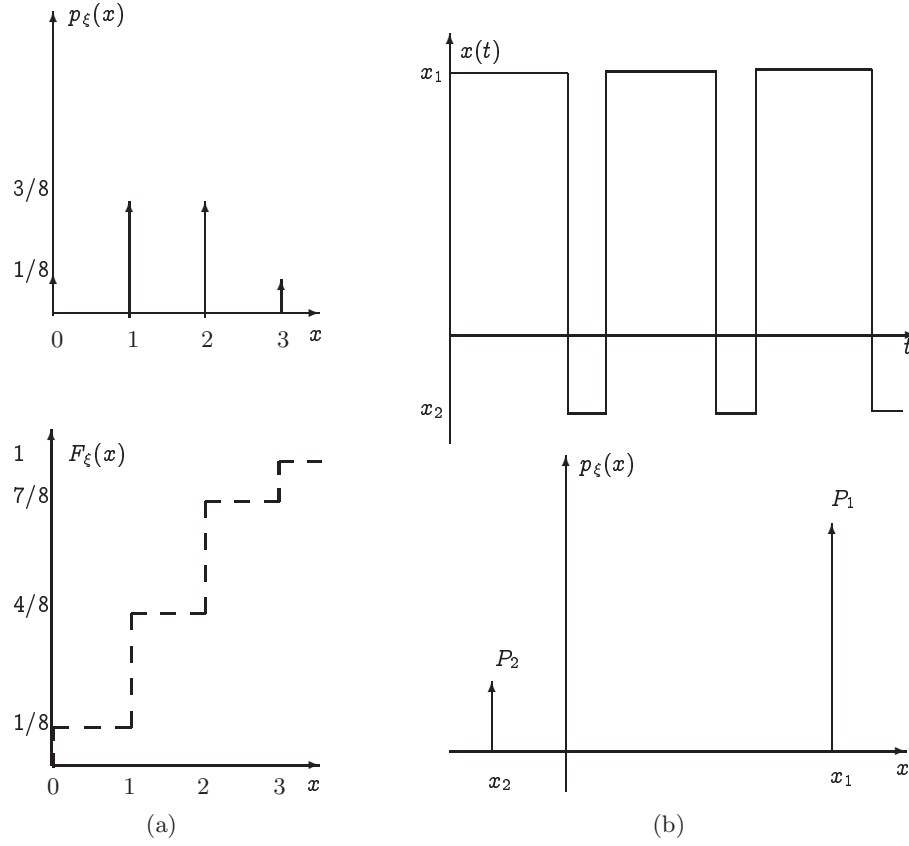


Figura 2.1. Ejemplo de FDP y función de distribución

**Definición 2.4.** Un conjunto de valores aleatorios  $\xi \in \mathbb{C}$ , determinados en un espacio probabilístico  $\{\Omega, \mathcal{B}, P\}$ , tal que su segundo momento sea finito,  $E\{\xi^2\} < \infty$ , conforma un espacio lineal normalizado de Hilbert  $L^2\{\Omega, \mathcal{B}, P\}$  provisto del producto escalar,

$$\langle \xi, \eta \rangle = E\{\xi \eta^*\}$$

con norma  $\|\xi\|_2 = (E\{\xi^2\})^{1/2}$ , mediante la cual se determina la distancia entre las respectivas variables aleatorias:

$$d(\xi, \eta) = \|\xi - \eta\|_2, \quad \xi, \eta \in L^2\{\Omega, \mathcal{B}, P\}$$

Además, toda variable aleatoria  $\xi$ , perteneciente a  $L^2\{\Omega, \mathcal{B}, P\}$ , es una variable aleatoria de Hilbert.

**Definición 2.5.** Las variables aleatorias de Hilbert,  $\xi$  y  $\eta$ , se definen ortogonales cuando se cumple que  $E\{\xi \eta^*\} = 0$ .

**Momento de tercer orden.** El momento central de tercer orden  $\mu_{3\xi}$  puede servir de criterio para estimar el grado de asimetría de la FDP con respecto al eje de abscisa que pasa por la coordenada con valor igual al primer momento de la variable aleatoria.

Frecuentemente, se emplea el *coeficiente de asimetría* definido por la siguiente relación de momentos de tercer orden:

$$\begin{aligned}\gamma_{1\xi} &= \frac{\mu_{3\xi}}{\sqrt{\mu_{2\xi}^3}} = \frac{\mu_{3\xi}}{\sigma_\xi^3} \\ &= \frac{m_{3\xi} - 3m_{1\xi}m_{2\xi} + 2m_{1\xi}^3}{\sigma_\xi^3}\end{aligned}\quad (2.15)$$

Cabe anotar, que cuando la FDP es simétrica, entonces  $\mu_{3\xi} = 0$  y, por lo tanto,  $\gamma_{1\xi} = 0$ . Por lo tanto, se puede juzgar que entre mayor sea el valor del coeficiente, mayor es la asimetría de la FDP.

**Momento de cuarto orden.** Se determina el *coeficiente de exceso*,

$$\begin{aligned}\gamma_{2\xi} &= \frac{\mu_{4\xi}}{\mu_{2\xi}^2} - 3 \\ &= \frac{m_{4\xi} - 4m_{2\xi}m_{1\xi} + 6m_{2\xi}m_{1\xi}^2 - 3m_{1\xi}^4}{\sigma_\xi^4} - 3,\end{aligned}\quad (2.16)$$

el cual es igual a 0 cuando corresponde a la FDP Gaussiana. El exceso compara el comportamiento del pico principal de cualquier FDP simétrica ( $\gamma_{1\xi} = 0$ ) con el pico de la densidad Gaussiana. Si resulta que  $\gamma_{2\xi} > 0$ , entonces el pico principal de la FDP en análisis es más grande y agudo que el pico de la densidad Gaussiana; asumiendo ambas FDP con igual valor medio y varianza.

El momento de orden  $n$  para la variable aleatoria  $|\xi|$  es su *momento absoluto de orden  $n$* , aunque los momentos de mayor orden se emplean con muy poca frecuencia. Sin embargo, es importante tener en cuenta que cualquiera que sea el orden del momento, su valor es único (en caso de existir) para una FDP dada. La afirmación inversa, según la cual a un conjunto dado de momentos la FDP se determina de forma única, no siempre es cierta (*problema de los momentos* [13]). Aunque en la mayoría de los casos prácticos, la afirmación se cumple en ambos sentidos.

Por último, cabe anotar que para algunas FDP, los momentos definidos tanto en (2.6) como en (2.9), puede ocurrir que no existan para ciertos valores de  $n$ .

**Cumulantes.** Basados en la definición dada, en la ecuación (2.7), se determina el valor medio de la función  $e^{j\omega\xi}$ , el cual se denomina *función característica*:

$$\Theta(j\omega) \triangleq E\{e^{j\omega\xi}\} = \int_{-\infty}^{\infty} p_\xi(x) e^{j\omega x} dx \quad (2.17)$$

que en esencia corresponde a la función de densidad espectral (1.18) de la función densidad de probabilidad, aunque con el signo invertido en el exponente.

Mediante la función característica también se pueden determinar los valores de los momentos iniciales, en particular, al diferenciar para  $\omega = 0$ ,

$$\begin{aligned} \left. \frac{d^k \Theta}{d\omega^k} \right|_{\omega=0} &= j^k \int_{-\infty}^{\infty} x^k e^{j\omega x} p_{\xi}(x) dx \bigg|_{\omega=0} = j^k \int_{-\infty}^{\infty} x^k p_{\xi}(x) dx \\ &= j^k m_{k\xi} \end{aligned}$$

de lo cual resulta que,

$$m_{k\xi} = \frac{1}{j^k} \left. \frac{d^k \Theta}{d\omega^k} \right|_{\omega=0}$$

Luego, los momentos (2.6) se pueden definir mediante el valor medio de la descomposición por la serie de MacLaurin:

$$\Theta(j\omega) = E \left\{ \sum_{k=0}^{\infty} \frac{1}{k!} \xi^k (j\omega)^k \right\} = 1 + \sum_{k=1}^{\infty} \frac{m_{k\xi}}{k!} (j\omega)^k$$

Los valores *semi-invariantes* o *cumulantes*  $m'_{k\xi}$  se definen del anterior promedio al evitar la influencia de la constante, para lo cual en (2.17) se halla el promedio de la función transformada en la forma  $\Psi(j\omega) = \ln \Theta(j\omega)$ :

$$\begin{cases} m'_{k\xi} &= j^{-k} \left( \frac{d^k \Psi(j\omega)}{d\omega^k} \right) \bigg|_{\omega=0} \\ \Psi(j\omega) &= \sum_{k=1}^{\infty} \frac{m'_{k\xi}}{k!} (j\omega)^k \end{cases} \quad (2.18)$$

Se observa de la expresión (2.18) que

$$\dot{\Psi}(j\omega) = \frac{\dot{\Theta}(j\omega)}{\Theta(j\omega)}, \quad \ddot{\Psi}(j\omega) = \frac{\Theta(j\omega)\ddot{\Theta}(j\omega) - (\dot{\Theta}(j\omega))^2}{\Theta^2(j\omega)}$$

De la definición de la función característica, se tiene que  $\Theta(0) = 1$ , luego,

$$\dot{\Psi}(0) = \dot{\Theta}(0), \quad \ddot{\Psi}(0) = \ddot{\Theta}(0) - (\dot{\Theta}(0))^2$$

Teniendo en cuenta la expresión (2.17), se obtienen los primeros dos momentos,

$$\begin{cases} \dot{\Theta}(0) &= jm_{1\xi}, \\ \ddot{\Theta}(0) &= -m_{2\xi}, \\ \ddot{\Psi}(0) &= -m_{2\xi} - (jm_{1\xi})^2 = -(m_{2\xi} - m_{1\xi}^2) = -\mu_{2\xi} \\ &= -\sigma_{\xi}^2 \end{cases} \quad (2.19)$$

Los cumulantes y momentos iniciales se relacionan mediante las siguientes expresiones:

$$\begin{aligned} m'_{1\xi} &= m_{1\xi}, m'_{2\xi} = m_{2\xi} - m_{1\xi}^2 = \sigma_\xi^2, m'_{3\xi} = m_{3\xi} - 3m_{1\xi}m_{2\xi} + 2m_{1\xi}^3, \dots \\ m_{1\xi} &= m'_{1\xi}, m_{2\xi} = m'_{2\xi} + m_{1\xi}^2, m_{3\xi} = m'_{3\xi} + 3m'_{1\xi}m'_{2\xi} + m_{1\xi}^3, \dots \end{aligned}$$

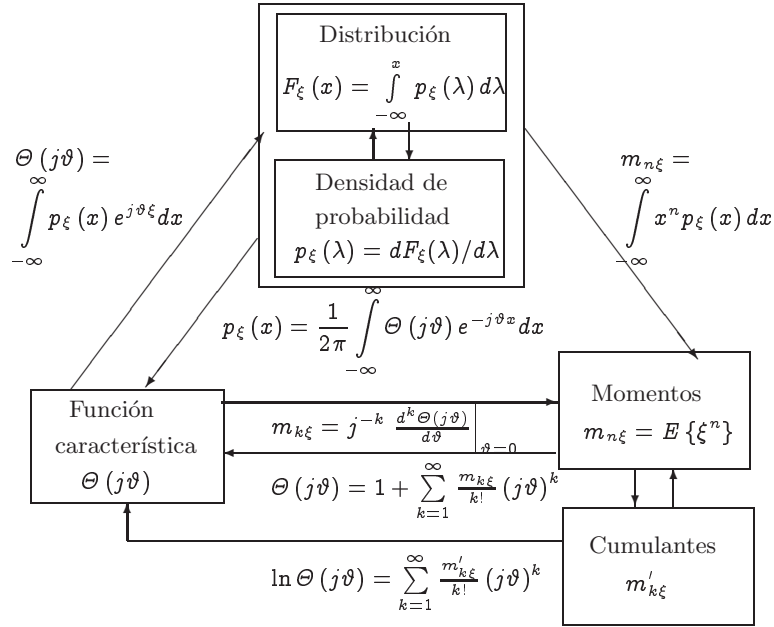
La asimetría y exceso se determinan a través de los cumulantes, respectivamente como,

$$\begin{cases} \gamma_{1\xi} &= m'_{3\xi} m'_{2\xi}^{-3/2}, \\ \gamma_{2\xi} &= m'_{4\xi} m'_{2\xi}^{-2} \end{cases} \quad (2.20)$$

Entre otros valores medios importantes están la *moda* ( $mo_\xi$ ), definida como el valor con mayor probabilidad de aparición de una variable aleatoria  $\xi$ , al cual le corresponde el máximo de la FDP y la *mediana* ( $me_\xi$ ), definida como el valor que satisface la condición:

$$P(\xi < me_\xi) = \int_{-\infty}^{me_\xi} p_\xi(x) dx = \int_{me_\xi}^{\infty} p_\xi(x) dx = \frac{1}{2}$$

Por cierto, si la FDP es una función simétrica, entonces,  $\gamma_{1\xi} = 0$ , además, se cumple que  $\text{sgn}(\gamma_{1\xi}) = \text{sgn}(me_\xi - mo_\xi)$ ,  $\forall \xi \neq 0$ .



**Figura 2.2.** Relación entre la distribución y las características de aleatoriedad

Las diversas relaciones de cálculo que existen entre la función de distribución (o la FDP), la función característica, los momentos y los cumulantes, se muestran en la Figura 2.2.



### 2.1.3. Funciones de probabilidad

Como se antes se dijo, en la conformación del espacio probabilístico de señales, las variables aleatorias, que tienen regularidad estadística, se pueden describir por medio de las funciones de probabilidad definidas en términos de  $\xi$ , en particular, las siguientes:

**Función de distribución acumulativa,  $F_\xi(x)$ .** Se define como la probabilidad de que la variable aleatoria continua tome valores menores que  $\xi = x_0$ :

$$F_\xi(x) = P\{\infty < \xi \leq x\} = P\{\xi \leq x\}$$

Mientras en el caso de variables aleatorias discreta  $\{x_i\} \in \xi$  con las respectivas probabilidades asociadas,  $P_i = P(\xi = x_i)$ , la distribución se expresa como:

$$F_\xi(x) = \sum_{i=1}^N P_i u(\xi - \xi_i)$$

La función de distribución acumulativa tiene las siguientes propiedades:

- (a).  $0 \leq F_\xi(x) \leq 1, \forall x$
- (b).  $F_\xi(x_m) \leq F_\xi(x_n), \forall x_m < x_n,$
- (c).  $F_\xi(-\infty) = 0$ , mientras  $F_\xi(\infty) = 1$ .

**Función densidad de probabilidad.** La variable aleatoria  $\xi$ , que tiene función de distribución absolutamente continua, esto es, la probabilidad de ocurrencia de cualquiera de su cantidad infinita de valores, sobre un intervalo tan angosto como se quiera, es infinitamente pequeña, entonces, existe una función  $p_\xi(x)$ , tal que

$$F_\xi(x) = \int_{-\infty}^x p_\xi(x) dx, \quad (2.21)$$

La función  $p_\xi(x)$  que cumple con (2.21) es la función densidad de probabilidad de la variable continua  $\xi$ , y corresponde al límite de la relación de la probabilidad de ocurrencia de los valores de la variable, dados en el intervalo  $[x, x + \Delta x]$ , sobre la longitud del mismo, cuando  $\Delta x \rightarrow 0$ :

$$p_\xi(x) = \lim_{\Delta x \rightarrow 0} P\{x \leq \xi < x + \Delta x\} / \Delta x$$

de lo anterior resulta que

$$P\{x \leq \xi < x + \Delta x\} = p_\xi(x) \Delta x + \varepsilon\{x\}, \quad \varepsilon\{x\} \geq 0 \quad (2.22)$$

siendo  $\varepsilon\{x\}$  una cantidad residual de mucho menor valor que  $\Delta x$ . La cantidad  $p(x) \Delta x$  corresponde al átomo de la FDP.

Las principales propiedades de la FDP son las siguientes:

- (a).  $p_{\xi}(x) \geq 0, \forall x \in \xi$
- (b).  $\int_{\xi} p_{\xi}(x) dx = F_{\xi}(\max \{x\}) = 1,$
- (c).  $\int_{x_1}^{x_2} p_{\xi}(x) dx = F_{\xi}(x_2) - F_{\xi}(x_1) = P\{x_1 \leq \xi < x_2\}$

Cabe anotar, que en el caso de que no exista la derivada (2.5), para la definición de  $p_{\xi}(x)$ , se recurre a la operación de límite.

De otra parte, para las variables aleatorias discretas se tiene que:

$$p_{\xi}(\xi) = \frac{d}{dx} \left( \sum_i P_i u(x - x_i) \right) = \sum_i P_i \delta(x - x_i),$$

La FDP puede ser discreto-análoga cuando se determina de forma combinada: sobre una malla de valores  $\{x_i \in \xi : i = 1, \dots, n\}$  con las respectivas probabilidades  $\{P_i : i = 1, \dots, n\}$  y sobre un intervalo  $\bar{x} \in \xi$ , para el cual se define un conjunto continuo de valores con densidad de probabilidad  $p_{\bar{x}}(x)$ , con lo cual la FDP de la variable discreto-análoga es la suma ponderada (donde los coeficientes de peso  $a, b$  deben cumplir la condición  $a + b = 1$ ):

$$p_{\xi}(x) = ap_{\bar{x}}(x) + b \sum_{i=1}^n P_i \delta(x - x_i)$$

**Ejemplo 2.6.** Un punto realiza la trayectoria oscilatoria  $x = a \sin \omega t$ . Hallar la FDP de la variable  $x$  en cualquier momento  $t$ , asumiendo, que la probabilidad de encontrarse el punto en el intervalo  $(x, x + dx)$  es proporcional a la longitud del intervalo  $dx$  e inversamente proporcional a la velocidad en el correspondiente momento de tiempo. Hallar la distribución integral y la probabilidad de que el punto se encuentra en el intervalo  $-a + b \leq x \leq a - c$ , ( $b < 2a; c < 2a$ ).

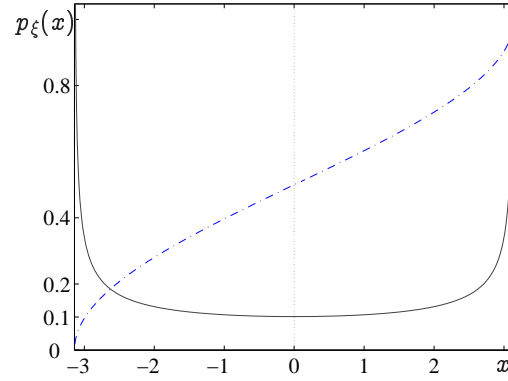
La condición de proporcionalidad para la probabilidad de encontrarse el punto en un intervalo dado se interpreta como,

$$P\{x \leq \xi \leq x + dx\} = k \frac{dx}{dx/dt} = k dt$$

por lo cual,  $p_{\xi}(x) = \frac{d}{dx} P\{\xi\} = k \frac{dt}{dx}$ . Por cuanto,  $dx/dt = a\omega \cos \omega t$ , entonces,

$$\frac{dt}{dx} = \frac{1}{a\omega \cos \omega t} = \frac{1}{a\omega \sqrt{1 - \sin^2 \omega t}} = \frac{1}{a\omega \sqrt{1 - \left(\frac{x}{a}\right)^2}} = \frac{1}{\omega \sqrt{a^2 - x^2}}$$

Asumiendo que  $|x| < a$ , la FDP se determina como,  $p_{\xi}(x) = k/(\omega \sqrt{a^2 - x^2})$ , donde la



**Figura 2.3.** Cálculo de la FDP — y distribución integral —.

constante  $k$  se calcula de la condición de unidad de área para cualquier FDP,

$$\int_{-a}^a \frac{k}{\omega \sqrt{a^2 - x^2}} dx = \frac{k}{\omega} \arccos \frac{x}{a} \Big|_{-a}^a = \frac{k}{\omega} \pi = 1$$

entonces,  $k = \omega/\pi$ , con lo que finalmente se obtiene que,  $p_{\xi}(x) = (\pi \sqrt{a^2 - x^2})^{-1}$ .

La Figura 2.3 muestra la función densidad de probabilidad obtenida (línea continua), en la cual se observa que su valor mínimo es  $\min\{p_{\xi}(x)\} = 1/\pi a$  que ocurre para  $x = 0$ , cuando la velocidad del punto es la máxima posible, mientras en los valores de  $x = \pm a$ , la densidad tiende a infinito, entonces, en esos puntos la FDP no tiene sentido.

La distribución integral (línea punteada), que existe en todo el intervalo de análisis, se calcula de la siguiente integral:

$$\begin{aligned} F_{\xi}(x) &= \int_{-a}^x p(x) dx = \int_{-a}^x \frac{dx}{\pi \sqrt{a^2 - x^2}} = \frac{1}{\pi} \arcsin \frac{x}{a} \Big|_{-a}^x \\ &= \frac{1}{\pi} \left( \arcsin \frac{x}{a} - \left( -\frac{\pi}{2} \right) \right) = \frac{1}{2} + \arcsin \frac{x}{a} \end{aligned}$$

**Variables aleatorias Gaussianas.** En la práctica, el modelo de FDP más empleado corresponde a la densidad *Normal* o *Gaussiana* (Figura 2.4(a)):

$$p(\xi) = \frac{1}{\sqrt{2\pi b}} \exp \left( -\frac{(\xi - a)^2}{2b^2} \right) \triangleq \mathcal{N}(a, b), \quad -\infty < \xi < \infty, \quad (2.23)$$

Los primeros momentos de la FDP Gaussiana, en concordancia con el modelo (2.23), se obtienen empleando la respectiva función característica,  $\Theta(j\omega) = \exp(j\omega a - b^2\omega^2/2)$ , mientras la función logarítmica característica es igual a,  $\Psi(j\omega) = j\omega a - b^2\omega^2/2$ .

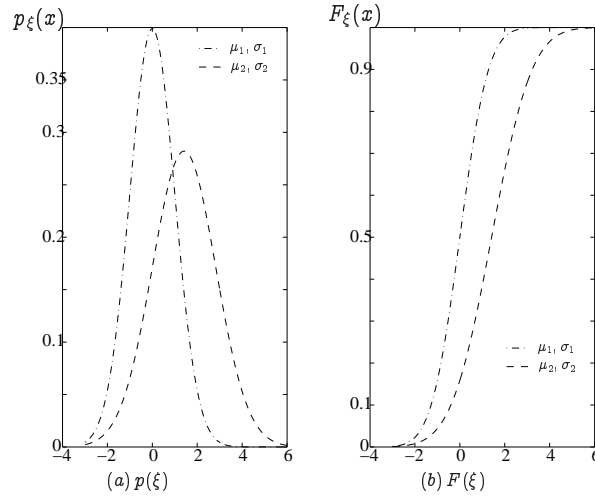
Por lo tanto, el momento inicial de primer orden, de acuerdo con (2.19), es igual a

$$m_{1\xi} = j^{-1} \dot{\Theta}(0) = j^{-1} \dot{\Psi}(0) = j^{-1}(ja) = a$$

Mientras, el momento central de segundo orden se calcula como,

$$\mu_{2\xi} = \sigma_{\xi}^2 = -\ddot{\Psi}(0) = -\frac{d^2}{d\omega^2} \left( j\omega a - \frac{b^2 \omega^2}{2} \right) \Big|_{\omega=0} = b^2$$

Así mismo, se calculan los momentos de tercer y cuarto orden, de acuerdo con la expresión (2.20):  $\mu_{3\xi} = 0$ ,  $\mu_{4\xi} = 3\mu_{2\xi}^2 = 3\sigma_{\xi}^4$ .



**Figura 2.4.** Distribución normal

La correspondiente función distributiva para un valor dado  $\xi_0$ , que se muestra en la Figura 2.4(b), se determina como:

$$F(\xi_0) = \int_{-\infty}^{\xi_0} \mathcal{N}(m_{1\xi}, \sigma_{\xi}) d\xi = \frac{1}{\sqrt{2\pi}\sigma_{\xi}} \int_{-\infty}^{\xi_0} \exp\left(-\frac{(\xi - m_{1\xi})^2}{2\sigma_{\xi}^2}\right) d\xi, \quad (2.24)$$

En la Figura 2.4 se muestran tanto la FDP como la distribución Gaussiana para diferentes valores de  $m_{1\xi_i}$  y  $\sigma_{\xi_i}$ ,  $i\{1, 2\}$ , en la que  $m_{1\xi_1} < m_{1\xi_2}$ , mientras,  $\sigma_{\xi_1} < \sigma_{\xi_2}$ .

En general, las siguientes son las razones por las cuales se emplea ampliamente la función densidad de probabilidad Gaussiana [14]:

1. Es un buen modelo matemático para la mayoría de los fenómenos aleatorios; es más, se puede demostrar que en muchos casos el modelo real converge al tipo Gaussiano (teorema del límite central).
2. La FDP normal es una de las pocas que puede describir fenómenos aleatorios compuestos por un número cualquiera de variables aleatorias.

3. Cualquier combinación lineal de variables aleatorias Gaussianas, también es Gaussiana.
4. Como se observa de (2.23), la FDP de Gauss se describe solo con dos parámetros: la media  $m_{1\xi}$  y la varianza  $\sigma_\xi^2$ .
5. El análisis matemático de los variables aleatorias muchas veces tiene solución solo para el caso Gaussiano.

Las principales propiedades de la FDP Gaussiana son las siguientes:

- (a). La curva normal es de forma simétrica con respecto a la media  $m_{1\xi}$ . En particular, a los valores de  $x_{a-} = m_{1\xi} - a$  y de  $x_{a+} = m_{1\xi} + a$  les corresponde un mismo valor de densidad.
- (b). El valor máximo de la curva está en  $\max(p_\xi(x)) = p_\xi(x = m_{1\xi}) = (\sigma_\xi \sqrt{2\pi})^{-1}$ , que es el valor de la variable aleatoria asignada e igual al primer momento inicial  $\xi$ . Dado  $\sigma_\xi = 1$ , entonces,  $\max(p_\xi(x)) = 0.39$ .
- (c). El valor de  $\sigma_\xi$  por la abscisa, que corresponde al punto de desdoblamiento de la curva, es  $0.6 \max(p_\xi(x))$ , con lo cual las  $2/3$  partes de todas las observaciones se distribuyen dentro del intervalo de los valores  $(m_{1\xi} - \sigma_\xi, m_{1\xi} + \sigma_\xi)$ .
- (d). Cualquier variable aleatoria con FDP Gaussiana, con alta probabilidad, toma valores cercanos a su esperanza, lo que se expresa por *la ley de los sigmas*:

$$P\{|\xi - m_{1\xi}| \geq k\sigma\} = \begin{cases} 0.3173, & k = 1 \\ 0.0455, & k = 2 \\ 0.0027, & k = 3 \end{cases} \quad (2.25)$$

Por cierto, tiene lugar la siguiente desigualdad:

$$P\{|\xi - m_{1\xi}| \geq k\sigma\} < 4/9k^2$$

que para  $k = 3$  implica que  $P\{|\xi - m_{1\xi}| \geq k\sigma\} < 4/81 \approx 0.05$ . En otras palabras, la probabilidad de desviación de la variable aleatoria del valor medio mayor a  $3\sigma_\xi$  es menor que 0.05.

- (e). En el caso de la ley de  $3\sigma$ , de la expresión (2.25) se pueden seleccionar los valores extremos que se consideran cumplen con la condición de estructura Gaussiana:

$$\xi_{\max} = m_{1\xi} + 3\sigma_\xi; \quad \xi_{\min} = m_{1\xi} - 3\sigma_\xi$$

La FDP *Gaussiana normalizada* se refiere al caso cuando la varianza  $\sigma_\xi^2$  es uno y la media es cero, esto es,  $(2\pi)^{-1/2} \exp(-\xi^2/2)$ . En este caso, la probabilidad de error para todos los

valores de la variable aleatoria  $\xi$  hasta  $x_0$  está dada por:

$$P\{\xi \leq x_0\} = \int_{-\infty}^{x_0} \mathcal{N}(0, 1) d\xi = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x_0} e^{-\frac{\xi^2}{2}} d\xi \triangleq \Phi(x_0), \quad (2.26)$$

Mediante el cambio de variables en (2.24), para la FDP Gaussiana se cumple que la probabilidad de ocurrencia de la variable aleatoria en el intervalo  $(x_1, x_2)$  es igual a:

$$P_\xi(x_1 \leq \xi < x_2) = \Phi\left(\frac{x_2 - m_{1\xi}}{\sigma_\xi}\right) - \Phi\left(\frac{x_1 - m_{1\xi}}{\sigma_\xi}\right)$$

En la práctica se tabula la función

$$\Phi_0(x_0) = \frac{1}{\sqrt{2\pi}} \int_0^{x_0} e^{-\frac{\zeta^2}{2}} d\zeta \quad (2.27)$$

conocida como la *Integral de Laplace* [1] y relacionada con  $\Phi(\xi)$  en (2.26) por la expresión,  $\Phi(\xi) = 1/2 + \Phi_0(\xi)$ . El cálculo de la integral, para valores pequeños de  $\xi$ , se realiza empleando la aproximación:

$$\Phi(\xi) = \frac{1}{2} + \frac{1}{\sqrt{2\pi}} \left( \xi - \frac{\xi^3}{2 \cdot 3} + \frac{\xi^5}{2! \cdot 2^2 \cdot 5} + \cdots + \frac{\xi^{2n+1}}{n! (2n+1) 2^n} + \cdots \right)$$

Otros ejemplos de modelos de FDP de amplio uso, para variables aleatorias continuas, son los siguientes:

### Densidad Gamma ( $\Gamma$ -densidad)

$$p_\xi(x; b, c) = \begin{cases} \frac{1}{b \Gamma(c)} \left(\frac{x-a}{b}\right)^{c-1} \exp\left(-\frac{x-a}{b}\right), & x \geq 0, b > 0, c \geq 0.5, x, a, b, c \in \mathbb{R} \\ 0, & \text{otros valores} \end{cases} \quad (2.28)$$

siendo  $x \in [0, \infty)$ , donde  $a$  se le conoce como el parámetro de posición, que generalmente se asume igual a cero,  $b > 0$  es de escala y  $c$  es el de forma.

La función Gamma  $\Gamma(n)$  ó función de Euler, que se define como,

$$\Gamma(n) = \int_0^\infty y^{n-1} \exp(-y) dy, \quad n > 0$$

Los momentos de orden  $k$  para la  $\Gamma$ -densidad están dados por la expresión

$$m_{k\xi} = \frac{c(c+1) \cdots (c+k-1)}{\lambda^k}, \quad \sigma_\xi^2 = c/\lambda^2$$

La  $\Gamma$ -densidad tiene importancia en la Teoría de Colas y describe el tiempo necesario para la aparición de  $c$  sucesos independientes que ocurren con igual intensidad  $\lambda$ .

---

**Ejemplo 2.7.** Un ferry comienza su ruta cada que se completan 9 automóviles, los cuales se consideran que llegan a la estación de manera independiente con una intensidad media de 6 unidades por hora. Determinar la probabilidad de que el tiempo entre los recorridos sucesivos del ferry sea menor que una hora.

Asumiendo que la variable aleatoria del tiempo entre las recorridos tiene  $\Gamma$ -densidad con parámetros:  $x = 1$ ,  $c = 9$ , y  $\lambda = 6$ , el cálculo de la respectiva probabilidad da como resultado  $P(1; 1/6, 9) = 0.153$ .

---

Como casos particulares de la  $\Gamma$ -densidad se consideran, entre otras, la FDP de *Erlang*, cuando  $c \in \mathbb{Z}^+$ , (entonces, se cumple que  $\Gamma(c) = (c-1)!$ ), la densidad  $\chi^2$  con  $n$  grados de libertad, cuando  $\lambda = 0.5$  y  $c = n/2$ . El caso especial, cuando  $c = 1.0$ , corresponde a la FDP *exponencial*, cuyo modelo se muestra en el ejemplo 2.8 y el cual se emplea en análisis de confiabilidad. Sin embargo, el parámetro  $\lambda$  asociado al tiempo medio entre fallas se asume constante, generando una fuerte restricción en el modelo exponencial.

Si se considera que  $\lambda(t) = \text{var}$  en el tiempo, particularmente,  $\lambda(t) = c/b((x-a)/b)^{c-1}$ , el modelo resultante de FDP corresponde a la densidad de *Weibull*.

$$p_{\xi}(x; a, b, c) = c/b((x-a)/b)^{c-1} \exp(-((x-a)/b)^c), \quad x \geq a, b > 0, c > 0 \quad (2.29)$$

La intensidad de fallas y la respectiva FDP toman diferentes formas variando el parámetro  $c$ . Particularmente, cuando  $c < 1$ , la FDP (2.29) es una curva monótona decreciente y con el transcurso del tiempo la intensidad de las fallas disminuye. Para  $c = 1$ , la intensidad de fallas es constante y la FDP de Weibull coincide con la exponencial, mientras para  $c > 1$  la densidad tiene forma de pico y la intensidad de fallas aumenta con el tiempo.

**Densidad de Student ( $t$ -densidad).** Corresponde a la FDP de la variable:

$$t = \frac{\xi_0}{\sqrt{1/n \left( \sum_{k=1}^n \xi_k^2 \right)}}$$

donde  $\xi_k$  son valores independientes, tales que  $\{\xi_k \in \mathcal{N}(0, 1) : k = 0, \dots, n\}$ . La densidad de la variable aleatoria  $t$  es igual a,

$$p_{\xi}(x; n) = \frac{\Gamma((n+1)/2)}{\sqrt{\pi n} \Gamma(n/2)} \left(1 + x^2/n\right)^{-(n+1)/2}, \quad -\infty < x < \infty \quad (2.30)$$

En variadas aplicaciones, se tiene que  $n \in \mathbb{N}$ , así, para  $n = 1$ , la función (2.30) corresponde a la FDP de *Cauchy* (que no tiene momentos ni de primer ni de segundo orden!):

$$p_{\xi}(x, \lambda) = \frac{1}{\pi} \frac{\lambda}{\lambda^2 + (x - n)}, \quad \lambda > 0$$

La forma de la  $t$ -densidad es similar a la curva de Gauss (para  $n \geq 40$  converge en la normal), no obstante, en lugar de depender de la media y la varianza, se expresa solamente en función del valor  $n$  ó *grados de libertad*. La principal aplicación de la  $t$ -densidad es la prueba de hipótesis para los valores medios.

**$\chi^2$ - densidad.** Sea la variable aleatoria  $\chi^2 = \sum_{k=1}^n \xi_k^2$ , conformada por los valores independientes  $\xi_k^2$ ,  $\{\xi_k \in \mathcal{N}(0, 1)\}$ , entonces la FDP con  $n$  grados de libertad:

$$p_{\xi}(x; n) = \begin{cases} \frac{1}{2^{\frac{n}{2}} \Gamma(\frac{n}{2})} x^{\frac{n}{2}-1} \exp(-\frac{x}{2}), & x > 0 \\ 0, & x \leq 0 \end{cases} \quad (2.31)$$

En la Figura 2.5, se muestran ejemplos para diferentes valores de  $n$  de la densidad  $\chi^2$ , la cual corresponde al caso particular de la  $\Gamma$ -densidad cuando  $b = 2$  y  $c = k/2$ . Esta FDP se determina sobre el dominio de valores  $0 \leq \xi < \infty$  y le corresponden los momentos de orden  $k$ :

$$m_{k\xi} = n(n+2) \cdots (n+2(k-1))$$

con varianza  $\sigma_{\chi^2}^2 = 2n$ .

Al aumentar el valor de  $n$ , la  $\chi^2$ -densidad tiende a la FDP normal, pero con menor velocidad que la  $t$ -densidad, por lo que para valores grandes de  $x$  y  $n$ , se demuestra que

$$F_{\chi^2(n)} \approx \Phi\left(\sqrt{2x} - \sqrt{2n-1}\right)$$

La  $\chi^2$ -densidad cumple el principio de aditividad,  $\chi^2(n_1 + n_2) = \chi^2(n_1) + \chi^2(n_2)$ .

El área principal de aplicación de esta FDP es la prueba de hipótesis. Cuando el número  $n$  de variables aleatorias a considerar es igual a 3:  $\chi^2 = \xi_1^2 + \xi_2^2 + \xi_3^2$ , asumiendo su estructura Gaussiana y su total independencia mutua, la respectiva FDP se denomina de *Maxwell*, mientras, para  $n = 2$ , corresponde a la FDP de *Rayleigh* y cuando es  $n = 1$ , se obtiene la FDP de la potencia de la variable original  $\xi$ . La densidad de Rayleigh también corresponde al caso particular de la FDP de Weibull, cuando  $a = 0$  y  $c = 2$ .

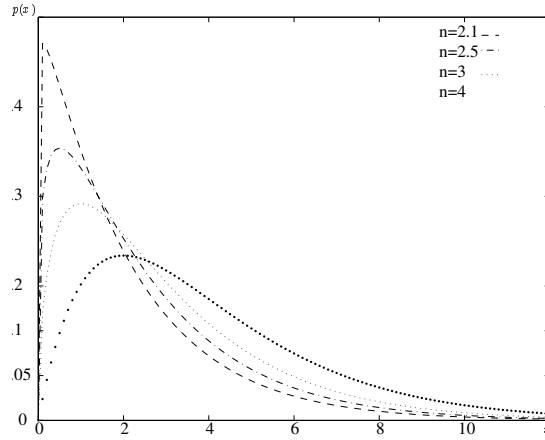


Figura 2.5. Densidad  $\chi^2$

**FDP de Fisher ( $F$ - densidad).** Corresponde a la FDP de la variable aleatoria

$$\xi = \left(\eta_{n_1/n_1}\right) / \left(\eta_{n_2/n_2}\right) \quad (2.32)$$



donde  $\eta_{n_i}$  son valores aleatorios independientes con  $\chi^2$ -distribución y  $n_i$ ,  $i = 1, 2$  grados de libertad. Un ejemplo particular, corresponde al caso cuando se tienen dos conjuntos de observaciones: el primero,  $\{x_i : i = 1, \dots, m_1\}$  y segundo,  $\{y_i : i = 1, \dots, m_2\}$ , ambas con FDP Gaussiana  $\mathcal{N}(m_{1\eta}, \sigma_\eta^2)$ ,  $\eta \in \{x, y\}$ . Entonces, la estadística

$$\frac{1}{m_1 - 1} \sum_{i=1}^{m_1} (x_i - \tilde{m}_{1x}) \bigg/ \frac{1}{m_2 - 1} \sum_{i=1}^{m_2} (y_i - \tilde{m}_{1y})$$

siendo  $\tilde{m}_{1\eta} = m_{1\eta}^{-1} \sum_{i=1}^{m_{1\eta}} \eta_i$ ,  $\eta \in \{x, y\}$ , le corresponde la  $F$ -densidad con  $(m_1, m_2)$  grados de libertad.

De otra parte, si se tiene la variable aleatoria  $\eta$  descrita por la  $F$ -densidad con  $(m_1, m_2)$  grados de libertad, entonces la variable aleatoria  $\xi = \frac{1}{2} \ln \eta$  le corresponde la  $z$ -densidad con  $(m_1, m_2)$  grados de libertad, para la cual se tienen los siguientes momentos:

$$m_{1\xi} = 0, \quad \sigma_\xi^2 = \frac{1}{2} \frac{m_1 + m_2}{m_1 m_2}$$

En general, la  $F$ -densidad se determina para el dominio  $0 \leq \xi < \infty$  y tiene aplicación en el análisis de varianza, la prueba múltiple de hipótesis y el análisis de regresión.

**Densidad logonormal.** Sea el par de variables aleatorias relacionadas por la transformación:  $\xi = \ln \eta$ , donde la variable aleatoria  $\xi$  es de estructura Gaussiana  $\mathcal{N}(m_{1\xi}, \sigma_\xi^2)$ , entonces se puede demostrar que la FDP para  $\eta$  tiene la forma

$$p_\eta(x) = \begin{cases} \frac{1}{x \sigma_\xi \sqrt{2\pi}} \exp\left(-\frac{(\ln x - m_{1\xi})^2}{2\sigma_\xi^2}\right), & x > 0 \\ 0, & x \leq 0 \end{cases}$$

con momentos,

$$m_{k\eta} = \exp\left(\frac{1}{2} k^2 \sigma_\xi^2 + k m_{1\xi}\right) \text{ y } \sigma_\eta^2 = \left(\exp(\sigma_\xi^2) - 1\right) \exp(\sigma_\xi^2 + 2m_{1\xi})$$

Si  $\eta$  es del tipo  $\mathcal{N}(0, 1)$ , entonces la variable  $\xi = \exp(\sigma_\xi \eta + m_{1\xi})$  tiene FDP logonormal con parámetros  $(m_{1\xi}, \sigma_\xi)$ . La FDP logonormal encuentra aplicación en tareas estadísticas relacionadas con mediciones experimentales de la física, geología, economía, biología, etc.

**Densidad uniforme.** Corresponde al caso en que todos los valores de la variable aleatoria tienen igual probabilidad de ocurrencia:

$$p_\xi(x) = \begin{cases} \frac{1}{b-a}, & x \in [a, b] \\ 0, & x \notin [a, b] \end{cases} \quad (2.33)$$

cuya función de distribución tiene la forma:

$$F_{\xi}(x) = \begin{cases} 0, & x < a \\ (x - a)/(b - a), & x \in [a, b] \\ 1, & x > b \end{cases}$$

La densidad uniforme le corresponden los momentos:

$$m_{k\xi} = \frac{b^{k+1} - a^{k+1}}{(b - a)(k + 1)}$$

con varianza  $\sigma_{\xi}^2 = (b - a)^2 / 12$ .

Al realizar la transformación lineal  $\eta = (\xi - a)/(b - a)$  se llega a la FDP uniforme en el intervalo  $[0, 1]$ . Esta propiedad explica su amplia aplicación en los métodos de simulación, por ejemplo el de Monte Carlo.

**Ejemplo 2.8.** Hallar los valores de la media y la varianza de las siguientes FDP:

**Exponencial.** Dada por el modelo de FDP (Figura 2.6a)

$$p_{\xi}(xi) = \begin{cases} \lambda e^{-\lambda x}, & x \geq 0 \\ 0, & x < 0 \end{cases} \quad (1)$$

a la cual le corresponden los valores de media y varianza, respectivamente,

$$m_{1\xi} = \lambda \int_0^{\infty} x e^{-\lambda x} dx = \frac{1}{\lambda}, \quad \sigma_{\xi}^2 = \lambda \int_0^{\infty} x^2 e^{-\lambda x} dx - \frac{1}{\lambda^2} = \frac{2}{\lambda^2} - \frac{1}{\lambda^2} = \frac{1}{\lambda^2}$$

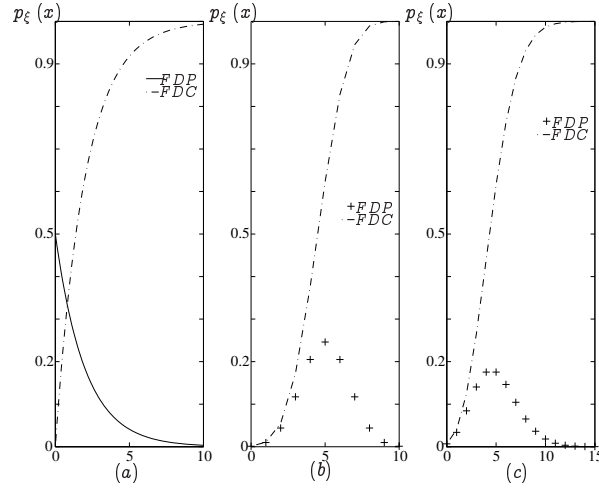
**Binomial.** Descrita por el modelo (Figura 2.6b):

$$p_k = P\{\xi = k\} = \begin{cases} 0, & k < 0 \\ C_n^k a^k (1 - a)^{n-k}, & k \leq n \\ 0, & k > n \end{cases} \quad (2)$$

con primer momento inicial:  $m_{1\xi} = \sum_{k=0}^n k C_n^k a^k (1 - a)^{n-k} = na$ , y segundo momento centralizado  $\sigma_{\xi}^2 = \sum_{k=0}^n k^2 C_n^k a^k (1 - a)^{n-k} = na(1 - a)$ .

**Poisson.** Para el caso asintótico de la FDP binomial, obtenido cuando  $n \rightarrow \infty$ , y  $a \rightarrow 0$ , siendo  $na = \lambda = \text{const.}$ , (el parámetro  $\lambda$  se denomina *valor promedio de ocurrencia* del evento), se puede demostrar que esta densidad se convierte en (Figura 2.6c):

$$p_k = \frac{\lambda^k}{k!} e^{-\lambda} u(k), \quad (3)$$



**Figura 2.6.** Ejemplos de distribuciones de probabilidad

A la densidad (3) le corresponden los siguientes momentos:

$$m_{1\xi} = \sum_{k=0}^{\infty} k \frac{\lambda^k}{k!} e^{-\lambda} = \lambda,$$

$$\sigma_{\xi}^2 = \sum_{k=0}^{\infty} k^2 \frac{\lambda^k}{k!} e^{-\lambda} = \sum_{k=0}^{\infty} k \frac{\lambda^k}{k!} e^{-\lambda} + \sum_{k=0}^{\infty} k(k-1) \frac{\lambda^k}{k!} e^{-\lambda} = \lambda$$

En la Figura 2.7 se muestran las relaciones que existen entre las diferentes densidades analizadas de probabilidad, en la cual algunos enlaces corresponden a simplificación o generalización de los modelos de aleatoriedad, o bien al empleo de la transformación  $\mathcal{G}\{\cdot\}$ . En líneas punteadas están indicadas las FDP discretas.

**Transformación de modelos de FDP.** Sea la variable aleatoria  $x \in \xi$  que se procesa en un dispositivo, cuya salida está dada por la relación  $y = f(x)$ ,  $y \in \eta$ . La transformación de la variable aleatoria se obtiene como:

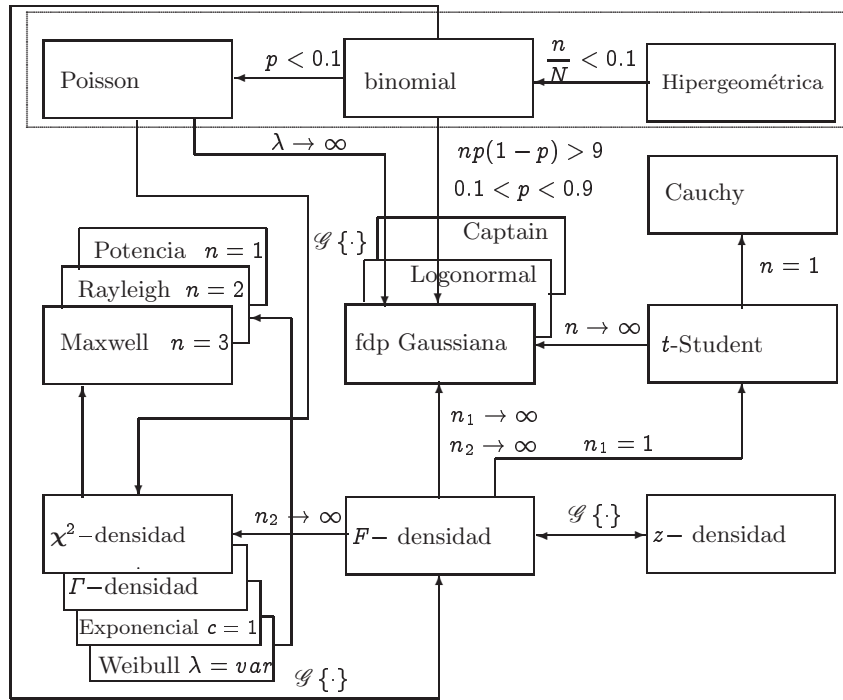
$$P\{\xi \leq x\} = P\{\xi \leq f^{-1}(y)\} = P\{\eta \leq y\}$$

además,

$$F_{\xi}(y) = P\{\eta \leq y\} = P\{f(x) \leq y\} = P\{x \leq f^{-1}(y)\} = F_{\xi}\{f^{-1}(y)\}$$

entonces, finalmente se obtiene

$$\frac{dF_{\eta}(y)}{dy} = \frac{dF_{\eta}(y)}{dx} \frac{dx}{dy} = \frac{dF_{\xi}(f^{-1}(y))}{dx} \frac{dx}{dy}.$$



**Figura 2.7.** Relación entre diferentes FDP.  $\mathcal{G}\{\cdot\}$  transformación funcional

Como es conocido, se debe cumplir que  $p(\xi) \geq 0$ , entonces, la definición de la FDP a la salida del dispositivo, expresada con respecto a las características de la variable aleatoria a la entrada, se ajusta de la siguiente manera:

$$p_{\eta}(y) = p_{\xi}(x) \frac{|dx|}{|dy|}, \quad (2.34)$$

donde el empleo de valores absolutos en (2.34) se hace para asegurar que en ningún momento la FDP sea negativa.

Cabe anotar, que el cálculo de los momentos después de la transformación se puede realizar sin necesidad de conocer la FDP de salida, en particular,

$$E\{(\eta - m_{1\eta})^n\} = \int_{-\infty}^{\infty} (y - m_{1\eta})^n p_{\eta}(y) dy = \int_{-\infty}^{\infty} (f(x) - m_{1\eta})^n p_{\xi}(x) dx \quad (2.35)$$

**Ejemplo 2.9.** Sea  $\eta = f(\xi) = 1 - e^{-\xi}$ ,  $0 \leq \xi \leq 2$ ,  $p(\xi) = ce^{-\xi}$ . Hallar el respectivo valor de  $c$  y la correspondiente FDP de salida,  $p(\eta)$ .

Por cuanto,  $\int_0^2 p(\xi) d\xi = c \int_0^2 e^{-\xi} d\xi = -ce^{-\xi} \Big|_0^2 = (-ce^{-2} + c) = 1$ .

Entonces,  $c = 1/0.865 = 1.156$ . Además,  $\xi = f^{-1}(\eta) = -\ln(1 - \eta)$ ,  $\frac{|d\eta|}{|d\xi|} = 1/(1 - \eta)$ ,

luego,  $p(\eta) = ce^{\ln(1-\eta)} 1 / (1 - \eta) = c = \frac{1}{\eta_{\max} - \eta_{\min}} = 1.156$ .

De lo anterior se deduce que la FDP  $p(\eta)$  corresponde al modelo uniforme.

**Ejemplo 2.10.** Hallar el valor medio y la varianza de la variable aleatoria de salida  $y \in \eta$ , dados los momentos  $m_{1\xi}$  y  $\mu_{2\xi}$  de la variable aleatoria entrada  $x \in \xi$ ; ambas variables relacionadas por la dependencia lineal  $y = ax + b$ .

De acuerdo con la expresión (2.35), se tiene que

$$\begin{aligned} E\{y\} &= \int_{-\infty}^{\infty} (ax + b) p_{\xi}(x) dx = a \int_{-\infty}^{\infty} x p_{\xi}(x) dx + b \int_{-\infty}^{\infty} p_{\xi}(x) dx \\ &= am_{1\xi} + b \\ E\{(y - m_{1\eta})^2\} &= \int_{-\infty}^{\infty} (ax + b - m_{1\eta})^2 p_{\eta}(x) dx = \int_{-\infty}^{\infty} (ax - b - am_{1\xi} - b)^2 p_{\xi}(x) dx \\ &= a^2 \int_{-\infty}^{\infty} (x - m_{1\xi})^2 p_{\xi}(x) dx = a^2 \sigma_{\xi}^2 \end{aligned}$$

**Descomposición de FDP.** En el análisis y transformación de variables aleatorias, puede ser de ayuda la descomposición ortogonal (1.4) de las diferentes FDP:

$$p_{\xi}(x) = \sum_{k=-\infty}^{\infty} x_k w_p(x) \phi_k(x) \quad (2.36)$$

siendo  $x_k$  los coeficientes de descomposición (1.7),  $\phi_k(x)$  la base ortogonal de descomposición y  $w_p(x)$  una función ventana de ajuste, que corresponde a la versión normalizada de la función de peso  $w(x)$ , definida en (1.9). En la práctica, en calidad de función ventana se toma una función de probabilidad conocida, por ejemplo la FDP normalizada,  $\mathcal{N}(0, 1)$ . Por cuanto el objetivo de la representación es la aproximación de las funciones de probabilidad, de tal manera que se pueda brindar la precisión deseada con el menor número de elementos de la serie en (2.36), entonces se emplean polinomios ortogonales, por ejemplo, los de *Chebyshev*, *Hermite*, *Laguerre*, *Legendre*, entre otros [3]

**Ejemplo 2.11.** Sea una FDP simétrica,  $p_{\xi}(x) = p_{\xi}(-x)$ , definida sobre el intervalo cerrado  $[-\pi \leq x \leq \pi]$ , e igual a  $p_{\xi}(x) = 0$  para todos los valores de  $x$  fuera del intervalo. Hallar la descomposición de la respectiva varianza por la serie de Fourier.

La descomposición por la serie (1.8), teniendo en cuenta la paridad de la FDP, se reduce solo a los términos reales,

$$p_{\xi}(x) = \frac{a_0}{2} + \sum_{n=1}^{\infty} a_n \cos nx$$

donde  $a_0$  se halla de la condición de unidad de área para  $p_\xi(x)$ ,

$$\int_{-\infty}^{\infty} \left( \frac{a_0}{2} + \sum_{n=1}^{\infty} a_n \cos nx \right) dx = 1$$

$$\text{entonces, } \frac{a_0}{2} x \Big|_{-\pi}^{\pi} + \sum_{n=1}^{\infty} \frac{a_n}{n} \sin nx \Big|_{-\pi}^{\pi} = 1.$$

Luego,  $a_0 = 1/\pi$ , y por lo tanto, la descomposición de la FDP, mediante la serie de Fourier, toma la forma

$$p_\xi(x) = \frac{1}{2\pi} + \sum_{n=1}^{\infty} a_n \cos nx$$

De la propiedad (2.13), se sabe que  $\sigma_\xi^2 = 2m_{1\xi}$ , con lo cual se obtiene el siguiente valor de varianza:

$$\sigma_\xi^2 = 2 \int_{-\infty}^{\infty} x^2 \left( \frac{1}{2\pi} + \sum_{n=1}^{\infty} a_n \cos nx \right) dx = \frac{\pi^2}{3} + 4\pi \sum_{n=1}^{\infty} (-1)^n \frac{a_n}{n^2}$$

En la práctica, se realizan las siguientes presunciones sobre la FDP a representar:

- (a). las funciones son unimodales, en el sentido en que presentan un solo máximo,
- (b). Las FDP presentan ramales por ambos lados del máximo, que caen suficientemente rápido hasta cero, en la medida en que aumenta el valor absoluto del argumento.

Un ejemplo corresponde a la descomposición mediante los polinomios de Hermite:

$$p_\xi(x) = w_p(x) \sum_{k=-\infty}^{\infty} \frac{x_k}{\sqrt{k!}} H_k(x) \quad (2.37)$$

para la cual, se tiene la ventana  $w(x) = e^{-x^2}$ , entonces  $w_p(x) = \mathcal{N}(0, 1)$ . Los coeficientes

$$x_k = \frac{1}{\sqrt{k!}} \int_{-\infty}^{\infty} p_\xi(x) H_k(x) dx = \frac{1}{\sqrt{k!}} E\{H_k(x)\}$$

son denominados *cuasimomentos* [15]. Donde los valores  $x_0 = 0$  y  $x_1 = x_2 = 0$ , debido a la condición de normalización de la función de densidad de probabilidad.

Si en la descomposición (2.37), la serie polinomial se limita hasta los primeros  $N$  términos, entonces se obtiene la *serie de Edworth*:

$$p_\xi(x) \approx w_p(x) \left( 1 + \sum_{n=3}^N \frac{x_k}{\sqrt{k!}} H_k(x) \right) = w_p(x) \left( 1 + \frac{\gamma_{1\xi}}{3!} H_3(x) + \frac{\gamma_{2\xi}}{4!} H_4(x) \right)$$

siendo  $\gamma_{i\xi}$ ,  $i = 1, 2$ , los momentos definidos en (2.15) y (2.16), respectivamente.

#### 2.1.4. Modelos de funciones de probabilidad con dimensión múltiple

Sea un fenómeno  $\xi$  descrito por  $n \geq 2$  variables aleatorias reales  $\{x_k \in \xi : k = 1, \dots, n\}$ , o *proceso aleatorio con dimensión múltiple*, para el cual los valores medios y momentos se encuentran empleando la integración múltiple por cada variable, incluyendo la función de densidad conjunta que también es de dimensión  $n$ .

**Modelos de probabilidad de dos dimensiones.** En el caso particular de dos dimensiones, la función de distribución acumulativa conjunta corresponde a la probabilidad de ocurrencia simultánea de las desigualdades  $x_1 \leq \lambda_1$ ,  $x_2 \leq \lambda_2$ :

$$F_\xi(\lambda_1, \lambda_2) = P\{x_1 \leq \lambda_1, x_2 \leq \lambda_2\} = \int_{-\infty}^{\lambda_1} \int_{-\infty}^{\lambda_2} p_\xi(x_1, x_2) dx_1 dx_2, \quad \lambda_1, \lambda_2 \in \mathbb{R} \quad (2.38)$$

Las siguientes propiedades se cumplen para las funciones de probabilidad de dos dimensiones:

- (a).  $0 \leq F_\xi(\lambda_1, \lambda_2) \leq 1$ . Por cierto,  $F_\xi(\lambda_1, -\infty) = F_\xi(-\infty, \lambda_2) = F_\xi(-\infty, -\infty) = 0$ , mientras, el otro caso extremo tiene valor  $F_\xi(\infty, \infty) = 1$
- (b).  $\frac{\partial}{\partial x_i} F_\xi(\lambda_1, \lambda_2) \geq 0$ ,  $i = 1, 2$ . De lo cual resulta que  $F_\xi(\lambda_1, x_2) \geq F_\xi(\lambda_2, x_2)$ , asumiendo que  $\lambda_2 > \lambda_1$ ;  $\lambda_1, \lambda_2 \in x_1$ , o bien,  $F_\xi(x_1, \lambda_1) \geq F_\xi(x_1, \lambda_2)$ ,  $\forall \lambda_2 > \lambda_1$ ;  $\lambda_1, \lambda_2 \in x_2$ .
- (c).  $p_\xi(x_1, x_2) = \frac{\partial^2}{\partial x_1 \partial x_2} F_\xi(x_1, x_2)$ ,
- (d).  $p_\xi(x_1, x_2) \geq 0$ ,

Las funciones de probabilidad de dimensión simple se expresan a través de las respectivas funciones de dos dimensiones, de la siguiente manera:

$$\begin{cases} F_\xi(\lambda_1) = \lim_{\lambda_2 \rightarrow \infty} F_\xi(\lambda_1, \lambda_2) = \int_{-\infty}^{\lambda_1} \int_{-\infty}^{\infty} p_\xi(x_1, x_2) dx_1 dx_2 \\ p_\xi(x_1) = \frac{\partial F_\xi(x_1)}{\partial x_1} = \int_{-\infty}^{\infty} p_\xi(x_1, x_2) dx_2, \end{cases} \quad (2.39)$$

De forma similar a lo obtenido en (2.22), para los valores suficientemente pequeños de  $\Delta_1 \in x_1$  y  $\Delta_2 \in x_2$ , tiene lugar la aproximación:

$$F_\xi(\lambda_1 \leq x_1 < \lambda_1 + \Delta_1, \lambda_2 \leq x_2 < \lambda_2 + \Delta_2) \approx p_\xi(x_1, x_2) \Delta_1 \Delta_2$$

Cuando las variables aleatorias  $x_1$  y  $x_2$ , que pertenecen al fenómeno  $\xi$ , son estadísticamente independientes, la condición necesaria y suficiente se expresa por medio de la relación

común con la FDP conjunta:

$$p_{\xi}(x_1, x_2) = p_{\xi}(x_1) p_{\xi}(x_2) \quad (2.40)$$

De otra parte, el nivel de dependencia estadística entre dos diferentes variables de un fenómeno aleatorio continuo aleatorio  $\xi$  está dado por la FDP condicional

$$\begin{cases} p_{\xi}(x_1 | x_2) = \frac{p_{\xi}(x_1, x_2)}{p_{\xi}(x_2)} = \frac{p_{\xi}(x_1, x_2)}{\int_{-\infty}^{\infty} p_{\xi}(x_1, x_2) dx_2} \\ p_{\xi}(x_1 | x_2) = \frac{p_{\xi}(x_1, x_2)}{p_{\xi}(x_1)} = \frac{p_{\xi}(x_1, x_2)}{\int_{-\infty}^{\infty} p_{\xi}(x_1, x_2) dx_1} \end{cases} \quad (2.41)$$

**Momentos de correlación.** Además de los valores de aleatoriedad definidos para variables con dimensión simple, en el caso particular de variables múltiples con dos dimensiones, se define el momento conjunto de primer orden [16]:

$$E\{g(x_1, x_2)\} \triangleq \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x_1, x_2) p_{\xi_1 \xi_2}(x_1, x_2) dx_1 dx_2 \quad (2.42)$$

siendo  $g(x_1, x_2)$  la *función generatriz* del momento conjunto.

La función generatriz más común corresponde al caso en que los momentos y los cumulantes de las variables aleatorias con dimensión múltiple se determinan como los coeficientes de descomposición en la serie exponencial de las respectivas funciones características o de sus logaritmos, por lo tanto similar al caso de análisis de una dimensión, los primeros coeficientes de descomposición son los más importantes y significativos.

Sea la función  $g(x_1, x_2)$ , que contiene las variables aleatorias  $\xi_1$  y  $\xi_2$ , cuyo valor medio se determina por (2.42). La descomposición de la función  $g(x_1, x_2)$  mediante la serie de Taylor en la vecindad de los puntos  $(m_{1\xi_1}, m_{1\xi_2})$ , donde se supone ocurre la mejor estimación de la función original, tiene la forma:

$$g(x_1, x_2) = g(m_{1\xi_2}, m_{1\xi_2}) + \frac{\partial g}{\partial x_1}(x_1 - m_{1\xi_1}) + \frac{\partial g}{\partial x_2}(x_2 - m_{1\xi_2}) + \dots$$

Si la FDP con dos dimensiones  $p_{\xi_1 \xi_2}(x_1, x_2)$  está concentrada en la cercanía de los puntos  $(\xi_1, \xi_2)$  y cambia suavemente en esta vecindad en proporción a  $(\sigma_{\xi_1}, \sigma_{\xi_2})$ , entonces la descomposición obtenida por Taylor, se puede limitar hasta los dos primeros términos:

$$E\{g(x_1, x_2)\} \simeq g(m_{1\xi_2}, m_{1\xi_2}) + \frac{1}{2} \left( \sigma_{\xi_1}^2 \frac{\partial^2 g}{\partial x_1^2} + 2\mu_{11} \frac{\partial^2 g}{\partial x_1 \partial x_2} + \sigma_{\xi_2}^2 \frac{\partial^2 g}{\partial x_2^2} \right) \quad (2.43)$$



siendo

$$\mu_{11} = E\{(\xi_1 - m_{1\xi_1})(\xi_2 - m_{1\xi_2})\} \triangleq K_{\xi_1\xi_2} \quad (2.44)$$

El valor del momento conjunto central de segundo orden  $\mu_{11}$  se denomina *momento* o *función de covarianza* conjunta entre las variables  $\xi_m$  y  $\xi_n$ , la cual se puede asimilar con el grado de dispersión o varianza mutua entre el par de variables aleatorias de dimensión simple. La expresión (2.43) muestra el papel importante que cumplen los momentos de menor orden, en particular, la esperanza, la varianza y el momento de correlación.

Se observa de (2.44) que la función generatriz es de la forma  $g(x_1, x_2) = x_1 x_2$ , por lo tanto, se define el momento inicial conjunto entre las variables  $\xi_m$  y  $\xi_n$  o *función de correlación*:

$$m_{11} = E\{\xi_1 \xi_2\} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x_1 x_2 p_{\xi_1, \xi_2}(x_1, x_2) dx_1 dx_2 \triangleq R_{\xi_1 \xi_2} \quad (2.45)$$

Al calcular los promedios conjuntos de dos variables, sin tener en cuenta sus respectivas escalas ni valores medios (obtenidos para el caso de dimensión simple), se emplea el *índice de correlación* o función de covarianza normalizada, que se define como:

$$\begin{aligned} \rho &\triangleq E\{(\xi_1 - m_{1\xi_1})/\sigma_{\xi_1} (\xi_2 - m_{1\xi_2})/\sigma_{\xi_2}\} \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{(x_1 - m_{1\xi_1})}{\sigma_{\xi_1}} \frac{(x_2 - m_{1\xi_2})}{\sigma_{\xi_2}} p_{\xi_1, \xi_2}(x_1, x_2) dx_1 dx_2 \end{aligned} \quad (2.46)$$

Entre las principales propiedades del coeficiente de correlación están las siguientes:

- (a).  $\rho = E\{\xi'_m \xi'_n\}$ , siendo  $\xi'_k = (\xi_k - m_{1\xi_k})/\sigma_{\xi_k}$ ,  $k = \{m, n\}$  la forma normalizada estadísticamente de la variable  $\xi_k$ , de tal manera que  $E\{\xi'_k\} = 0$  y  $\sigma_{\xi'_k}^2 = 1$ .
- (b).  $-1 \leq \rho \leq 1$ , por cuanto,

$$\begin{aligned} E\{(\xi'_m \pm \xi'_n)^2\} &= E\{\xi'^2_m \pm 2\xi'_m \xi'_n + \xi'^2_n\} = 1 \pm 2\rho + 1 \\ &= 2(1 \pm \rho) \geq 0 \end{aligned}$$

A partir de la definición 2.5, se tiene que un valor 0 para la función de correlación (2.45), implica que ambas variables,  $\xi_m$  y  $\xi_n$ , son ortogonales entre sí. Si la respectiva función de covarianza (2.44) es cero, entonces se dice que las variables  $\xi_m$  y  $\xi_n$  son linealmente independientes. Por cierto, de la definición (2.44) se deduce que dos variables aleatorias ortogonales cumplen la condición de independencia lineal si sus respectivos valores medios son iguales a cero. Por último, se tiene un tipo de independencia más fuerte cuando las variables aleatorias  $\xi_m$  y  $\xi_n$  son estadísticamente independientes, entonces se cumple que  $\rho = E\{\xi'_m \xi'_n\} = 0$ . En este caso, la FDP conjunta corresponde a la expresión (2.40), de la cual resulta que,  $E\{\xi_m, \xi_n\} = m_{1\xi_m} m_{1\xi_n}$ , por lo que la independencia estadística implica

la independencia lineal. El caso contrario no siempre es cierto, por cuanto la independencia estadística es una condición más fuerte que la independencia lineal.

---

**Ejemplo 2.12.** Hallar los momentos  $\mu_{11}$  y  $\sigma_{\xi \pm \eta}^2$  de las variables dependientes  $\eta$  y  $\xi$ .

Teniendo en cuenta la expresión (2.44), se tiene

$$\begin{aligned}\mu_{11} &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x_1 - m_{1\xi_1})(x_2 - m_{1\xi_2}) p_{\xi_1, \xi_2}(x_1, x_2) dx_1 dx_2 \\ &= E\{\xi_1 \xi_2\} - m_{1\xi_1} E\{\xi_2\} - m_{1\xi_2} E\{\xi_1\} + m_{1\xi_1} m_{1\xi_2} \\ &= E\{\xi_1, \xi_2\} - m_{1\xi_1} m_{1\xi_2}\end{aligned}$$

De la definición de la varianza se tiene,

$$\begin{aligned}\sigma_{\xi \pm \eta}^2 &= E\{(\xi \pm \eta)^2\} - (E\{\xi \pm \eta\})^2 = E\{\xi^2 \pm 2\xi\eta + \eta^2\} - (E\{\xi\} \pm E\{\eta\})^2 \\ &= E\{\xi^2\} \pm E\{2\xi\eta\} + E\{\eta^2\} - E^2\{\xi\} \mp 2E\{\xi\}E\{\eta\} - E^2\{\eta\} \\ &= E\{\xi^2\} - E^2\{\xi\} + E\{\eta^2\} - E^2\{\eta\} \pm E\{2\xi\eta\} \\ &= \sigma_{\xi_1}^2 + \sigma_{\xi_2}^2 \pm E\{2\xi\eta\}\end{aligned}$$

El anterior resultado se puede generalizar para la suma de  $m$  variables dependientes:

$$\sigma^2 \left( \sum_{l=1}^m \xi_l \right) = \sum_{l=1}^m \sum_{n=1}^m K_{nl}$$


---

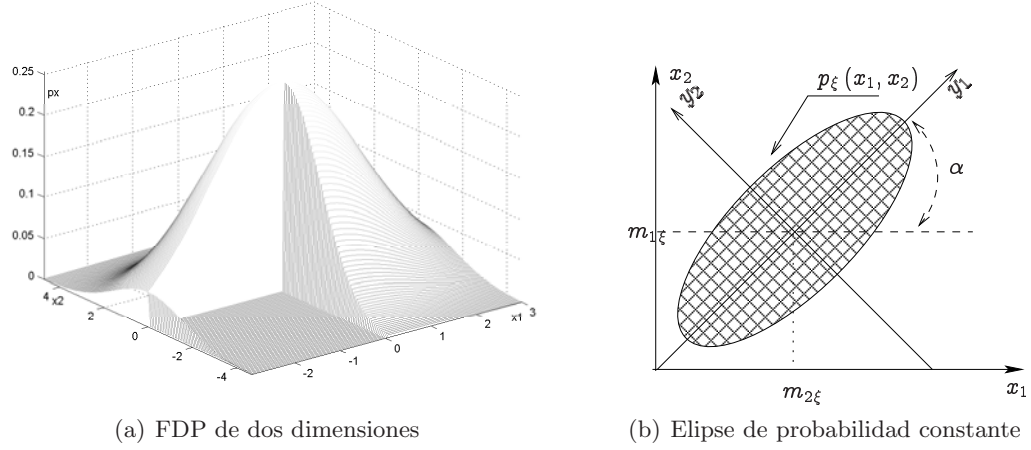
**Variables conjuntas Gaussianas.** Por definición, dos variables aleatorias  $\xi_1$  y  $\xi_2$  se denominan *conjuntamente Gaussianas* si su FDP conjunta tiene la forma:

$$p_{\xi_1 \xi_2}(x_1, x_2) = k \exp \left( - \left( ax_1^2 + bx_1x_2 + cx_2^2 + dx_1 + ex_2 \right) \right), \quad k \in \mathbb{R}$$

tal que la forma cuadrática de la exponencial,  $(ax_1^2 + bx_1x_2 + cx_2^2 + dx_1 + ex_2) \geq 0$ ,  $\forall x_1$  y  $x_2$ . La FDP Gaussiana conjunta expresada en función de sus momentos toma la forma:

$$\begin{aligned}p_{\xi_1 \xi_2}(x_1, x_2) &= \frac{1}{2\pi\sigma_{\xi_1}\sigma_{\xi_2}\sqrt{1-\rho^2}} \times \\ &\times \exp \left( \frac{-\sigma_{\xi_2}^2(x_1 - m_{1\xi})^2 + 2\rho\sigma_{\xi_1}\sigma_{\xi_2}(x_1 - m_{1\xi})(x_2 - m_{2\xi}) - \sigma_{\xi_1}^2(x_2 - m_{2\xi})^2}{2\sigma_{\xi_1}^2\sigma_{\xi_2}^2(1-\rho^2)} \right)\end{aligned}\tag{2.47}$$

Las siguientes son las propiedades de la FDP Gaussiana de 2 dimensiones (Figura 2.8(a)), las cuales se pueden generalizar a casos con dimensión múltiple de mayor orden.



**Figura 2.8.** Distribución de Gauss de dos dimensiones

- (a). Si dos variables aleatorias  $\xi_1$  y  $\xi_2$  no están correlacionadas, esto es,  $\rho = 0$ , entonces de (2.47) resulta que su FDP conjunta es igual a la multiplicación de la densidad de probabilidad de cada una de las variables:  $p_{\xi_1 \xi_2}(x_1, x_2) = p_{\xi_1}(x_1)p_{\xi_2}(x_2)$ . De acuerdo con (2.40), las variables son estadísticamente independientes entre sí.
- (b). Dos variables aleatorias,  $\xi_1$  y  $\xi_2$ , conjuntas Gaussianas dependientes (con valor de correlación  $\rho \neq 0$ ), siempre se pueden expresar en forma de dos nuevas variables aleatorias,  $\eta_1$  y  $\eta_2$ , que no tengan correlación (independientes), usando la siguiente transformación lineal (elipse de probabilidad constante de la Figura 2.8(b)):

$$\eta_1 = (\xi_1 - m_{1\xi_1}) \cos \alpha + (\xi_2 - m_{1\xi_2}) \sin \alpha, \quad \eta_2 = -(\xi_1 - m_{1\xi_1}) \sin \alpha + (\xi_2 - m_{1\xi_2}) \cos \alpha$$

- (c). Si dos variables aleatorias son conjuntamente Gaussianas, entonces, cada una de estas también tiene FDP Gaussiana. La afirmación inversa sólo se cumple cuando las variables aleatorias son independientes entre sí.

**Ejemplo 2.13.** Sea la FDP conjunta Gaussiana (2.47) de un fenómeno aleatorio descrito por las variables aleatorias  $(x, y)$  con parámetros  $m_\xi, \sigma_\xi^2$ ,  $\xi \in x, y$  y  $\rho_{xy}$ . Determinar las densidades marginales  $p_\xi(x), p_\xi(y)$ , así como las FDP condicionales  $p_\xi(y|x)$  y  $p_\xi(x|y)$ .

La densidad marginal se halla a partir de la expresión (2.39), en la cual para la FDP conjunta Gaussiana (2.47), se puede realizar el siguiente cambio de variables:

$$(x - m_x) / \sqrt{2}\sigma_x = u, \quad (y - m_y) / \sqrt{2}\sigma_y = v$$

con lo cual se obtiene

$$p_\xi(x) = \frac{1}{\sqrt{2\pi}\sigma_x\sqrt{1-\rho_{xy}^2}} \exp\left(-\frac{u^2}{1-\rho_{xy}^2}\right) \int_{-\infty}^{\infty} \exp\left(-\frac{1}{1-\rho_{xy}^2}v^2 + \frac{2\rho_{xy}u}{1-\rho_{xy}^2}v\right) dv$$

Teniendo en cuenta que

$$\int_{-\infty}^{\infty} \exp(-p^2 x^2 \pm qx) dx = \frac{\sqrt{\pi}}{p} \exp\left(\frac{q^2}{4p^2}\right)$$

se halla la densidad respectiva marginal

$$p_{\xi}(x) = \frac{1}{\sqrt{2\pi}\sigma_x} e^{-u^2} = \frac{1}{\sqrt{2\pi}\sigma_x} \exp\left(-\frac{(x - m_{1x})^2}{2\sigma_x^2}\right)$$

De la expresión anterior, es clara la estructura Gaussiana del variable  $x$  con media  $m_{1x}$  y varianza  $\sigma_x^2$ . De forma similar, se encuentra la densidad marginal alterna por la variable  $y$ , que también resulta ser de estructura Gaussiana con momentos  $m_{1y}, \sigma_y^2$ . Así mismo, de (2.41) se hallan las correspondientes FDP condicionales:

$$p(y|x) = \frac{1}{\sigma_y \sqrt{2\pi} \sqrt{1 - \rho_{xy}^2}} \exp\left(-\frac{1}{2(1 - \rho_{xy}^2)} \left(y - m_{1y} - \rho_{xy} \frac{\sigma_y}{\sigma_x} (x - m_{1x})\right)^2\right)$$

$$p(x|y) = \frac{1}{\sigma_x \sqrt{2\pi} \sqrt{1 - \rho_{xy}^2}} \exp\left(-\frac{1}{2(1 - \rho_{xy}^2)} \left(x - m_{1x} - \rho_{xy} \frac{\sigma_x}{\sigma_y} (y - m_{1y})\right)^2\right)$$

las cuales corresponden a las respectivas variables Gaussianas con parámetros

$$m_{y|x} = m_y + \rho_{xy} \frac{\sigma_y}{\sigma_x} (x - m_{1x}), \sigma_{y|x} = \sigma_y \sqrt{1 - \rho_{xy}^2}$$

$$m_{x|y} = m_x + \rho_{xy} \frac{\sigma_x}{\sigma_y} (y - m_{1y}), \sigma_{x|y} = \sigma_x \sqrt{1 - \rho_{xy}^2}$$

**Transformación de FDP en modelos con dos dimensiones.** Sea el par de variables aleatorias  $\xi_i, i = 1, 2$ , definidas por la relación mutua:  $\eta_i = g_i(\xi_1, \xi_2)$ ,  $i = 1, 2$ , siendo  $g_i, i = 1, 2$  funciones determinísticas conocidas. Por cuanto, la distribución conjunta se da por la relación:

$$F_{\eta_1 \eta_2}(y_1, y_2) = P\{\eta_1 < y_1, \eta_2 < y_2\} = P\{g_1(\xi_1, \xi_2) < y_1, g_2(\xi_1, \xi_2) < y_2\}$$

entonces

$$F_{\eta_1 \eta_2}(y_1, y_2) = \int \int_D p_{\eta_1 \eta_2}(x_1, x_2) dx_1 \quad (2.48)$$

donde el área de integración  $D$  se determina por las desigualdades:

$$g_1(x_1, x_2) < y_1, \quad g_2(x_1, x_2) < y_2$$

La FDP transformada se obtiene diferenciando la expresión (2.48):

$$p_{\eta_1 \eta_2}(y_1, y_2) = \partial^2 F_{\eta_1 \eta_2} / \partial y_1 \partial y_2 \quad (2.49)$$

La densidad obtenida (2.49) se puede calcular directamente, sin necesidad de determinar inicialmente la función  $F_{\eta_1\eta_2}(y_1, y_2)$ , empleando las respectivas expresiones de relación inversa, esto es,  $\xi_i = g_i^{-1}(\eta_1, \eta_2)$ ,  $i = 1, 2$ . Particularmente,

$$p_{\eta_1\eta_2}(y_1, y_2) = p_{\eta_1\eta_2}(g_1^{-1}(y_1, y_2), g_2^{-1}(y_1, y_2)) |J(y_1, y_2)| \quad (2.50)$$

donde

$$J(y_1, y_2) = \frac{\partial(x_1, x_2)}{\partial(y_1, y_2)} = \begin{bmatrix} \partial g_1^{-1}/\partial y_1 & \partial g_1^{-1}/\partial y_2 \\ \partial g_2^{-1}/\partial y_1 & \partial g_2^{-1}/\partial y_2 \end{bmatrix} = \left( \frac{\partial(y_1, y_2)}{\partial(x_1, x_2)} \right)^{-1} \quad (2.51)$$

es el Jacobiano de la transformación de las variables aleatorias  $\{\xi_i\}$  en las variables,  $\{\eta_i\}$ ,  $i = 1, 2$ , respectivamente.

**Ejemplo 2.14.** Sea la FDP conjunta  $p_{\xi_1\xi_2}(x_1, x_2)$  correspondiente a las variables aleatorias  $\xi_1, \xi_2$ . Dados los coeficientes constantes  $a, b, c, d$ , hallar la FDP de 2 dimensiones  $p_{\eta_1\eta_2}(y_1, y_2)$ , para las variables aleatorias  $\eta_1, \eta_2$ , transformadas mediante las relaciones

$$\eta_1 = a\xi_1 + b\xi_2, \quad \eta_2 = c\xi_1 + d\xi_2$$

La solución exige que el determinante del sistema, compuesto por los coeficientes  $a, b, c, d$ , sea diferente de cero. Entonces el sistema compuesto por el par de ecuaciones algebraicas lineales  $y_1 = ax_1 + bx_2$ ,  $y_2 = cx_1 + dx_2$  tiene la única solución:

$$x_1 = a_1y_1 + b_1y_2, \quad x_2 = c_1y_1 + d_1y_2$$

en la cual, los coeficientes  $a_1, b_1, c_1, d_1$  se expresan a través de el respectivo conjunto  $a, b, c, d$ . En este caso, el Jacobiano de la transformación (2.51) tiene la forma:

$$\begin{aligned} J(x_1, x_2) &= \frac{\partial(x_1, x_2)}{\partial(y_1, y_2)} = \left( \frac{\partial(y_1, y_2)}{\partial(x_1, x_2)} \right)^{-1} = \frac{1}{\begin{bmatrix} \partial y_1/\partial x_1 & \partial y_1/\partial x_2 \\ \partial y_2/\partial x_1 & \partial y_2/\partial x_2 \end{bmatrix}} = \frac{1}{\begin{bmatrix} a & b \\ c & d \end{bmatrix}} \\ &= \frac{1}{ad - bc} \end{aligned}$$

Por lo tanto, en correspondencia con (2.50), se obtiene

$$p_{\eta_1\eta_2}(y_1, y_2) = \frac{1}{|ad - bc|} p_{\eta_1\eta_2}(a_1y_1 + b_1y_2, c_1y_1 + d_1y_2)$$

**Ejemplo 2.15.** Sean dos variables Gaussianas independientes  $x$  e  $y$  con valores medios cero,  $m_{1x} = m_{1y} = 0$ , e igual varianza,  $\sigma_x^2 = \sigma_y^2 = \sigma^2$ , que corresponden a la representación cartesiana de un conjunto de puntos. Hallar las densidades de las variables transformadas  $r$  y  $\varphi$ , que representan al módulo y fase de las coordenadas polares y relacionadas con las cartesianas en la forma,  $x = r \cos \varphi$ ,  $y = r \sin \varphi$ .

Debido a la independencia estadística de las variables cartesianas, entonces la FDP conjunta corresponde a la multiplicación de las FDP marginales originales (ver ec. (2.40)),

$$p(x, y) = p(x) p(y) = \frac{1}{2\pi\sigma^2} \exp\left(-\frac{(x^2 + y^2)}{2\sigma^2}\right)$$

La FDP conjunta de dos dimensiones de las variables  $r$  y  $\varphi$ , a partir de la ec. (2.50), son iguales a  $p(r, \varphi) = p(x(r, \varphi), y(r, \varphi)) J(r, \varphi)$ , siendo

$$J(r, \varphi) = \begin{bmatrix} \frac{\partial x}{\partial r} & \frac{\partial x}{\partial \varphi} \\ \frac{\partial y}{\partial r} & \frac{\partial y}{\partial \varphi} \end{bmatrix} = \begin{bmatrix} \cos \varphi & -r \sin \varphi \\ \sin \varphi & r \cos \varphi \end{bmatrix} = r$$

Cada una de la distribuciones marginales de las variables transformadas se halla integrando su FDP conjunta de dos dimensiones por la variable aleatoria alterna,

$$\begin{aligned} p(r) &= \int_{-\pi}^{\pi} p(r, \varphi) d\varphi = \frac{r}{2\pi\sigma^2} \exp\left(-\frac{r^2}{2\sigma^2}\right) \int_{-\pi}^{\pi} d\varphi \\ &= \frac{r}{\sigma^2} \exp\left(-\frac{r^2}{2\sigma^2}\right), \quad r > 0 \end{aligned} \quad (1)$$

La FDP obtenida para el módulo corresponde a la densidad de Rayleigh, cuando  $n = 2$  en la expresión (2.31) para la  $\chi^2$ -densidad. De manera similar, se halla la FDP de la fase,

$$\begin{aligned} p(\varphi) &= \int_0^{\infty} p(r, \varphi) dr = \frac{1}{2\pi} \int_0^{\infty} \frac{r}{\sigma^2} \exp\left(-\frac{r^2}{2\sigma^2}\right) dr \\ &= \frac{1}{2\pi}, \quad -\pi < \varphi < \pi \end{aligned}$$

A partir de las relaciones obtenidas para la transformación para FDP de dos dimensiones, se pueden hallar los momentos conjuntos de las funciones  $\zeta_i = g_i(\xi, \eta)$ ,  $i = 1, 2$ , de manera directa tomando como base la densidad conjunta  $p_{\xi\eta}(x, y)$  de las respectivas variables aleatorias  $\xi$  y  $\eta$ . En este caso la esperanza matemática, la varianza y la función de correlación se determinan de forma correspondiente por las siguientes expresiones:

$$m_{1\zeta_i} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g_i(x, y) p_{\xi\eta}(x, y) dx dy, \quad i = 1, 2 \quad (2.52a)$$

$$\sigma_{\zeta_i}^2 = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (g_i(x, y) - m_{1\zeta_i})^2 p_{\xi\eta}(x, y) dx dy, \quad i = 1, 2 \quad (2.52b)$$

$$R_{\zeta_1\zeta_2} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (g_1(x, y) - m_{1\zeta_1})(g_2(x, y) - m_{1\zeta_2}) p_{\xi\eta}(x, y) dx dy \quad (2.52c)$$

En particular, de (2.52a), (2.52b) y (2.52c) se tienen las siguientes propiedades:

(a). Sea  $c$  una variable no aleatoria, entonces

$$E\{c\} = c, \quad E\{c\xi\} = cE\{\xi\}$$

(b). Para cualquier par de variables aleatorias  $\xi$  y  $\eta$ , se cumple que

$$E\{\xi\eta\} = E\{\xi\}E\{\eta\} - R_{xy}$$

Por cierto, si se cumple que ambas variables  $\xi$  y  $\eta$  son no correlacionadas, esto es, tienen correlación nula,  $R_{xy} = 0$ , entonces,  $E\{\xi\eta\} = E\{\xi\}E\{\eta\}$ .

(c). Sea  $c$  una variable no aleatoria, entonces

$$\sigma_c^2 = 0, \quad \sigma_{c\xi}^2 = c^2\sigma_\xi^2, \quad \sigma_{c\xi} = |c|\sigma_\xi$$

(d). Para cualquier par de variables aleatorias  $\xi$  y  $\eta$ ,

$$\sigma_{\xi \pm \eta}^2 = \sigma_\xi^2 + \sigma_\eta^2 \pm 2R_{xy}(\tau)$$

Si ambas variables aleatorias  $\xi$  y  $\eta$  tienen correlación cero, entonces se cumple que,

$$\sigma_{\xi \pm \eta}^2 = \sigma_\xi^2 + \sigma_\eta^2$$

(e). Para cualquier par de variables aleatorias independientes  $\xi$  y  $\eta$

$$\sigma_{\xi\eta}^2 = \sigma_\xi^2\sigma_\eta^2 + m_{1\xi}^2\sigma_\eta^2 + m_{1\eta}^2\sigma_\xi^2$$

**Ejemplo 2.16.** Hallar la media, la varianza y función de correlación de las variables aleatorias  $\{\zeta_i, i = 1, 2\}$ , expresadas en función de las variables  $\xi$  y  $\eta$ , en la forma aditiva,

$$\zeta_i = k_{1i}\xi + k_{2i}\eta, \quad k_{ni} = \text{const.}, \quad n, i = 1, 2.$$

Basados en las propiedades, expuestas anteriormente, se obtiene

$$\begin{aligned} E\{\zeta_i\} &= E\{k_{1i}\xi + k_{2i}\eta\} = k_{1i}E\{\xi\} + k_{2i}E\{\eta\} \\ \sigma_{\zeta_i}^2 &= \sigma_{k_{1i}\xi + k_{2i}\eta}^2 = k_{1i}^2\sigma_\xi^2 + k_{2i}^2\sigma_\eta^2 + 2R_{xy} \\ &= k_{1i}^2\sigma_\xi^2 + k_{2i}^2\sigma_\eta^2 + 2k_{1i}k_{2i}\sigma_\xi\sigma_\eta\rho_{\xi\eta} \\ R_{\zeta_1\zeta_2} &= E\{(\zeta_1 - m_{1\zeta_1})(\zeta_2 - m_{1\zeta_2})\} = E\{\zeta_1\zeta_2\} - E\{\zeta_1\}E\{\zeta_2\} \\ &= E\{(k_{11}\xi + k_{21}\eta)(k_{12}\xi + k_{22}\eta)\} - (k_{11}E\{\xi\} + k_{21}E\{\eta\})(k_{12}E\{\xi\} + k_{22}E\{\eta\}) \\ &= k_{11}k_{12}\sigma_\xi^2 + k_{21}k_{22}\sigma_\eta^2 + (k_{21}k_{12} + k_{11}k_{22})R_{\xi\eta} \\ &= k_{11}k_{12}\sigma_\xi^2 + k_{21}k_{22}\sigma_\eta^2 + \sigma_\xi\sigma_\eta\rho_{\xi\eta} \end{aligned}$$

**Modelos de probabilidad de  $n$  dimensiones.** Los diversos modelos de FDP se generalizan para el caso de un número finito de dimensiones  $n$ , correspondiente a las variables aleatorias  $x_1, \dots, x_n$  del proceso  $\xi$ . En este caso, la generalización de (2.38) para la distribución con dimensión  $F_\xi(x_1, \dots, x_n)$  corresponde a la probabilidad de que conjuntamente

cada variable aleatoria  $x_k$  tome el respectivo valor menor que  $\lambda_k$ ,  $k = 1, \dots, n$ , esto es,

$$\begin{aligned} P\{x_1 \leq \lambda_1, x_2 \leq \lambda_2, \dots, x_n \leq \lambda_n\} &= F_\xi(\lambda_1, \lambda_2, \dots, \lambda_n) \\ &= \int_{-\infty}^{\lambda_1} \int_{-\infty}^{\lambda_2} \dots \int_{-\infty}^{\lambda_n} p_\xi(x_1, x_2, \dots, x_n) dx_1 dx_2 \dots dx_n \end{aligned}$$

La generalización de las propiedades obtenidas para las FDP de 2 dimensiones, extendida al caso de variables aleatorias con mayor dimensión, corresponde a las expresiones:

(a).  $0 \leq F_\xi(\lambda_1, \lambda_2, \dots, \lambda_n) \leq 1$ . Donde

$$\begin{aligned} F_\xi(-\infty, x_2, \dots, x_n) &= F_\xi(x_1, -\infty, \dots, x_n) = F_\xi(x_1, x_2, \dots, -\infty) = 0 \\ F_\xi(\infty, \dots, \infty, x_{m+1}, \dots, x_n) &= F_\xi(x_{m+1}, \dots, x_n), \quad \forall m \leq n \end{aligned}$$

(b).  $\frac{\partial}{\partial x_i} F_\xi(\lambda_1, \lambda_2, \dots, \lambda_n) \geq 0$ ,  $i = 1, 2, \dots, n$ . De lo cual, resulta que

$$F_\xi(x_1, \dots, \lambda_{k1}, \dots, x_n) \geq F_\xi(x_1, \dots, \lambda_{k2}, \dots, x_n), \quad \forall \lambda_{k1} \geq \lambda_{k2}, \lambda_{k1}, \lambda_{k2} \in x_k$$

(c). Si existe la respectiva función densidad de probabilidad, entonces

$$p_\xi(x_1, x_2, \dots, x_n) = \frac{\partial^n F_\xi(x_1, x_2, \dots, x_n)}{\partial x_1 \partial x_2 \dots \partial x_n} \geq 0$$

Además, se cumple que,

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} p_\xi(x_1, x_2, \dots, x_n) dx_1 dx_2 \dots dx_n = 1$$

Las variables aleatorias con dimensión  $n$  se denominan *mutuamente no correlacionadas* si para las respectivas funciones de probabilidad se cumple que

$$F_\xi(x_1, \dots, x_n) = \prod_{k=1}^n F_\xi(x_k) \quad (2.53a)$$

$$p_\xi(x_1, x_2, \dots, x_n) = \prod_{k=1}^n p_\xi(x_k) \quad (2.53b)$$

Las expresiones (2.53a) y (2.53b) generalizan la relación (2.40). No obstante, aunque se puede afirmar que de un conjunto de variables aleatorias no correlacionadas con dimensión  $n$  resulta la independencia estadística por todas las posibles parejas de variables aleatorias  $x_i$  y  $x_j$ ,  $i, j = 1, \dots, n$ , la afirmación inversa, en forma general, no es cierta [12], excepto para la FDP Gaussiana [15].



El conjunto de variables aleatorias  $\{x_i : i = 1, \dots, n\}$  se puede entender como las coordenadas puntuales o componentes del vector aleatorio  $\boldsymbol{\xi}$  sobre un espacio con dimensión  $n$ . En este caso, para cumplir la condición de independencia estadística de cualquier par de vectores aleatorios  $\{x_i : i = 1, \dots, n\} \in \boldsymbol{\xi}$  y  $\{y_i : i = 1, \dots, n\} \in \boldsymbol{\eta}$  es necesario y suficiente que la densidad conjunta de sus componentes sea igual al producto de las FDP de las componentes de cada uno de los vectores:

$$p(\boldsymbol{\xi}, \boldsymbol{\eta}) = p_{\boldsymbol{\xi}\boldsymbol{\eta}}(x_1, \dots, x_n, y_1, \dots, y_n) = p_{\boldsymbol{\xi}}(x_1, \dots, x_n) p_{\boldsymbol{\eta}}(y_1, \dots, y_n)$$

El valor medio de la variable  $x_k$ , para el caso de modelos de FDP con dimensión múltiple se generaliza en la forma,

$$m_{1\xi}(x_k) = \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} x_k p_{\xi}(x_1, \dots, x_n) dx_1 \dots dx_n$$

La respectiva covarianza del par de variables  $x_k$  y  $x_l$ ,  $\forall k, l = 1, \dots, n$ , es igual a,

$$\mu_{11}(x_k, x_l) = \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} (x_k - m_{1\xi}(x_k))(x_l - m_{1\xi}(x_l)) p_{\xi}(x_1, \dots, x_n) dx_1 \dots dx_n,$$

Cabe anotar, que para el caso de  $k = l$ , el momento  $\mu_{11}(x_k, x_k)$  coincide con la varianza de  $x_k$ , esto es,  $\mu_{11}(x_k, x_k) = \sigma_{x_k}^2$ .

En forma general, los momentos centralizados conjuntos de cualquier orden se determinan por la expresión

$$\begin{aligned} \mu_{k_1 k_2 \dots k_n}(x_1, \dots, x_n) \\ = \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} (x_1 - m_{1\xi}(x_1))^{k_1} \dots (x_n - m_{1\xi}(x_n))^{k_n} p_{\xi}(x_1, \dots, x_n) dx_1 \dots dx_n \end{aligned}$$

siendo  $k_l \in \mathbb{N}$ ,  $l = 1, \dots, n$ .

Usualmente, cuando se tiene una cantidad de variables  $n > 2$ , es preferible emplear la *matriz de covarianza*, definida como

$$\mathbf{K}_{n \times n} \triangleq \begin{bmatrix} K_{x_1 x_1} & K_{x_1 x_2} & \dots & K_{x_1 x_n} \\ K_{x_2 x_1} & K_{x_2 x_2} & \dots & K_{x_2 x_n} \\ \dots & \dots & \dots & \dots \\ K_{x_n x_1} & K_{x_n x_2} & \dots & K_{x_n x_n} \end{bmatrix} \quad (2.54)$$

De la definición (2.44) para la función de covarianza resulta que  $K_{x_k x_l} = K_{x_l x_k}$ , además

$K_{x_k x_k} = \sigma_{x_k}^2$ , por lo tanto es frecuente el empleo de la siguiente matriz triangular

$$\mathbf{K} = \begin{bmatrix} \sigma_{x_1}^2 & K_{x_1 x_2} & \cdots & K_{x_1 x_n} \\ & \sigma_{x_2}^2 & \cdots & K_{x_2 x_n} \\ & & \cdots & \cdots \\ & & & \sigma_{x_n}^2 \end{bmatrix}$$

Si el conjunto de variables  $\{x_k : 1, \dots, n\}$  cumple las condiciones de independencia estadística, entonces los momentos cruzados centrales de primer orden son iguales a cero, con lo cual se obtiene la siguiente matriz diagonal:

$$\mathbf{K} = \begin{bmatrix} \sigma_{x_1}^2 & 0 & 0 & 0 \\ & \sigma_{x_2}^2 & \cdots & 0 \\ & & \cdots & \cdots \\ & & & \sigma_{x_n}^2 \end{bmatrix}$$

**Modelos Gaussianos con  $n$  dimensiones.** Los modelos de distribución Gaussiana conjunta con dimensión  $n > 2$ , es preferible, expresarlos en forma matricial:

$$p_{\xi}(\mathbf{x}) = \frac{1}{(2\pi)^{n/2} |\mathbf{K}|^{1/2}} \exp \left( -\frac{1}{2} (\mathbf{x} - \mathbf{m}_{1\mathbf{x}})^{\top} \mathbf{K}^{-1} (\mathbf{x} - \mathbf{m}_{1\mathbf{x}}) \right) \quad (2.55)$$

siendo  $\mathbf{x}$  y  $\mathbf{m}$  los vectores columna conformados por los elementos  $\{x_k : k = 1, \dots, n\}$  y  $\{m_{1x_k} : k = 1, \dots, n\}$ , respectivamente. Además,  $|\mathbf{K}| = \det(\mathbf{K})$ .

De la expresión (2.55) se observa, que cuando el conjunto de variables aleatorias Gaussianas  $\{x_k\}$  tiene covarianza nula, esto es,  $\{K_{x_k x_l} = 0 : \forall k \neq l; k, l = 1, \dots, n\}$ , entonces, las variables cumplen la condición de independencia estadística dada en (2.53b).

En forma general, se puede demostrar que las siguientes son las respectivas relaciones recursivas en los momentos iniciales y centralizados para la FDP Gaussiana :

$$m_{k\xi} = \frac{k(k-1)}{2} \int_0^{\sigma_{\xi}^2} m_{k-2,\xi} (m_{1\xi}, \sigma_{\xi}^2) d\sigma_{\xi}^2 + m_{1\xi}^k \quad (2.56a)$$

$$\mu_k = \frac{k(k-1)}{2} \int_0^{\sigma_{\xi}^2} \mu_{k-2} d\sigma_{\xi}^2 = \begin{cases} 1 \times 3 \times 5 \dots (k-1) \sigma_{\xi}^k, & k \in \text{par} \\ 0, & k \in \text{impar} \end{cases} \quad (2.56b)$$

### 2.1.5. Medidas de información en variables aleatorias

**Información en señales discretas.** La medida de información contenida en un proceso aleatorio, que toma los valores discretos  $\{x_n : n = 1, \dots, N\}$ , cumple las restricciones:

1. La cantidad de información debe ser un valor aditivo, esto es, en dos procesos aleatorios independientes la medida es igual a la suma de las cantidades en cada proceso.

2. La cantidad de información en un proceso no aleatorio (determinístico) es cero.
3. La cantidad de información no debe depender de algún factor subjetivo.

De lo anterior, la información propia de un evento  $x_i$ , también denominado *símbolo*, se asume mediante la medida,  $I(x_i) = -\log_2 p(x_i)$ , donde  $p(x_i)$  es la probabilidad a priori de aparición del valor  $x_i$ . El logaritmo se toma en base 2 y, en consecuencia, las unidades de información son *bits* (en adelante el índice de la base se omitirá).

Otras medidas de información se determinan por las siguientes relaciones:

- Información propia condicional de la variable  $x_m$ , conocida la variable aleatoria  $y_n$ ,

$$I(x_m|y_n) = -\log p(x_m|y_n)$$

- Información mutua de dos variables o información de  $y_n$  con respecto a  $x_m$ ,

$$I(x_m; y_n) = \log p(x_m|y_n) \quad (2.57)$$

- Información propia del evento conjunto  $(x_m, y_n)$ ,

$$I(x_m, y_n) = -\log p(x_m, y_n)$$

- Cantidad media de información, presente en  $x_n$  como elemento del alfabeto  $\{x_n\}$ ,

$$I(X; y_m) = \sum_{n=1}^N I(x_n; y_m) p(x_n|y_m) = \sum_{n=1}^N p(x_n|y_m) \log \frac{p(x_n|y_m)}{p(x_n)}$$

- Cantidad media de información mutua sobre los elementos correspondientes a un alfabeto  $Y = \{y_m : m = 1 \dots, M\}$  para un valor determinado de  $x_n$ ,

$$I(x_n; Y) = \sum_{m=1}^M I(x_n; y_m) p(y_m|x_n) = \sum_{m=1}^M p(y_m|x_n) \log \frac{p(y_m|x_n)}{p(y_m)}$$

- Valor medio de la cantidad completa de información mutua entre conjuntos  $(Y, X)$ ,

$$I(X; Y) = \sum_{n=1}^N \sum_{m=1}^M p(x_n, y_m) \log \frac{p(x_n|y_m)}{p(x_n)} = \sum_{m=1}^M p(y_m) I(X; y_m)$$

Si  $x_n$  se relaciona estadísticamente con los valores  $y_m$  y  $z_k$ ,  $k = 1, \dots, K$ , conocidas las FDP condicionales  $p(x_n, y_m, z_k)$ , la información condicional mutua se determina por,

$$I(x_m; y_n|z_k) \log \frac{p(x_m|y_n, z_k)}{p(x_m|z_k)} = \log \frac{p(x_m; y_n|z_k)}{p(x_m|z_k) p(y_n|z_k)}$$

Las siguientes relaciones son válidas para las cantidades puntuales de información:

$$\begin{aligned}
I(x_n; y_m) &= I(y_m; x_n), I(x_n; y_m) \leq I(x_n), I(x_n; y_m) \leq I(y_m) \\
I(x_n; y_m, z_k) &= I(x_n; y_m) + I(x_n; z_k | y_m) = I(x_n; z_k) + I(x_n; y_m | z_k) \\
I(x_n; y_m) &= I(x_n) - I(x_n | y_m) = I(y_m) - I(y_m | x_n) = I(x_n) + I(y_m) - I(x_n, y_m) \\
I(x_n, y_m) &= I(x_n) + I(y_m) - I(x_n; y_m) \\
I(X; y_m) &\geq 0, I(x_n, Y) \geq 0 \\
I(X, Y) &= I(Y, X) \geq 0, \\
I(X; Y, Z) &= I(X, Y) + I(X; Z | Y) \\
I(Y, Z; X) &= I(Y; X) + I(Z; X | Y)
\end{aligned}$$

De (2.8), se define el valor medio de la información propia para un alfabeto discreto:

$$I(X) = E\{I(x_n)\} = \sum_{n=1}^N p(x_n) I(x_n) = - \sum_{n=1}^N p(x_n) \log p(x_n) \triangleq H(X) \quad (2.58)$$

El valor  $H(X)$  se denomina *entropía* de la variable aleatoria  $X$  y corresponde a la medida cuantitativa de su incertidumbre. La entropía es la principal característica de un fuente de información: entre más alto sea el valor de la entropía mayor es la información contenida por el conjunto  $X$ .

La entropía tiene las siguientes propiedades:

- (a). *Definida positiva*:  $H(X) \geq 0$ ,
- (b). *Aditividad*. Sea  $H(Y|X)$  la entropía condicional de los valores del conjunto  $Y$ , dado un conjunto de eventos  $X$ , entonces, la entropía  $H(Y, X)$  del conjunto de eventos conjuntos  $X$  y  $Y$  se define como:

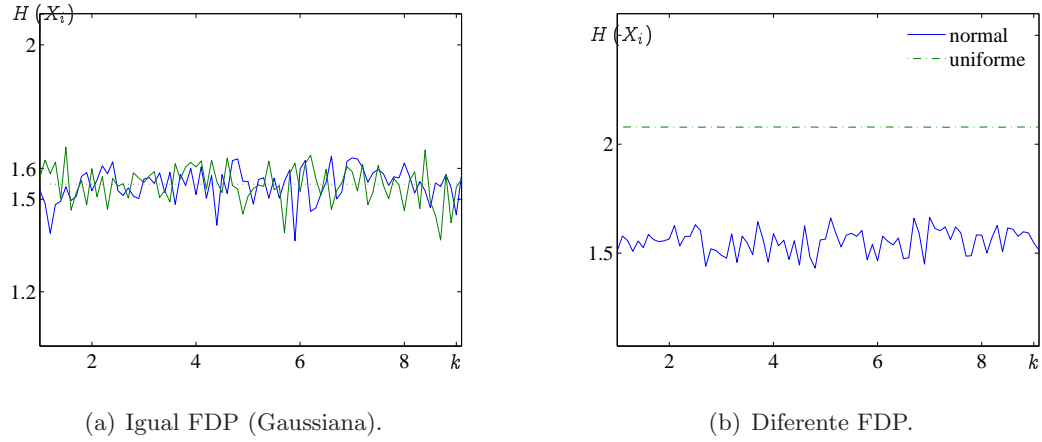
$$H(Y, X) = H(X) + H(Y|X) = H(Y) + H(X|Y) \quad (2.59)$$

Por cierto, si los conjuntos  $X$  e  $Y$  son independientes, entonces  $H(Y|X) = H(Y)$ , por lo tanto, se cumple que  $H(Y, X) = H(X) + H(Y)$ .

- (c). *Valor acotado*. Si el conjunto de valores  $X = \{x_n : n = 1, \dots, N\}$ , la máxima entropía contenida está acotada por el valor  $H(X) \leq \lg N$ , donde la igualdad tiene lugar cuando todos los valores  $x_n$  son equiprobables e independientes estadísticamente.

**Ejemplo 2.17.** Desarrollar un programa que simule dos observaciones con longitud  $N$ , para las variables aleatorias  $X_i \in \{x_l : l = 1, \dots, L\}$ ,  $i = 1, 2$ , que permita analizar el comportamiento de sus respectivos valores de entropía para los siguientes casos:

- Ambas variables tienen FDP Gaussiana, con valor medio cero. Estimar las entropías para valores de relación de las varianzas:  $\sigma_{X_1}^2 = k\sigma_{X_2}^2$ ,  $k = 1, \dots, 10$ .
- La variable  $X_1$  tiene FDP Gaussiana, mientras la variable  $X_2$  se genera con FDP uniforme; ambas con los mismos valores medio y de varianza.



**Figura 2.9.** Entropía en función de  $k = \sigma_{X_2}^2 / \sigma_{X_1}^2$ ,  $N = 4096$  y  $L = 8$ .

En la Figura 2.9(a), se muestran los resultados para el caso de generación de ambas variables Gaussianas con media  $m_{1X_1} = m_{1X_2} = 0$ , pero  $\sigma_{X_1}^2 = k\sigma_{X_2}^2$ . Se observa que la estimación de la entropía converge a un mismo valor para ambos casos de análisis. La Figura 2.9(b) corresponde al caso de diferentes FDP (Gaussiana y uniforme) y muestra que el valor de la entropía, para dos sucesiones aleatorias con iguales valores medio y de varianza, depende de la estructura de aleatoriedad.

**Ejemplo 2.18.** Demostrar que para dos sucesos con probabilidades respectivas,  $p_1$  y  $p_2$ , la mayor entropía se obtiene para  $p_1 = p_2$ .

Al reemplazar en la definición de entropía (2.58),  $p_2 = 1 - p_1$ , se tiene que

$$H = -(p_1 \ln p_1 + p_2 \ln p_2) = -(p_1 \ln p_1 + (1 - p_1) \ln (1 - p_1))$$

El máximo valor de  $p_1$  se obtiene al diferenciar e igualar a cero la anterior ecuación,

$$\frac{dH}{dp_1} = - \left( p_1 \frac{1}{p_1} + \ln p_1 - \ln (1 - p_1) \right) = 0$$

Como resultado, al resolver la ecuación se obtiene que  $p_1 = 0.5$ , luego,  $p_1 = p_2$ .

La entropía condicional se define como:

$$H(Y|X) = - \sum_{n=1}^N \sum_{m=1}^M p(x_n, y_m) \log p(y_m|x_n) = -p(x_n) \sum_{m=1}^M p(y_m|x_n) \log p(y_m|x_n)$$

para la cual se cumple que,  $0 \leq H(Y|X) \leq H(Y)$ .

El valor medio de la información mutua  $I(X; Y)$  se relaciona con la entropía como:

$$\begin{aligned} I(X; Y) &= H(X) - H(X|Y) = H(Y) - H(Y|X) = H(X) + H(Y) - H(X, Y) \\ I(X; Y) &\leq H(X), \quad I(X; Y) \leq H(Y) \end{aligned}$$

La entropía es una medida adecuada de la incertidumbre contendida en la FDP y su uso es preferible en aquellos casos cuando la densidad es asimétrica, con muchos picos de concentración, que es cuando los momentos pierden su efecto interpretativo del proceso.

De otra parte, la medida de incertidumbre  $H(X)$  depende de la relación estadística entre los elementos  $\{x_i \in X\}$ , por lo que se puede definir qué tanto de la entropía máxima potencial no se contiene en el conjunto  $X$ , a través de la medida de relación denominada *redundancia*:

$$R = \frac{H_{\max} - H(X)}{H_{\max}} = \frac{\log N - H(X)}{\log N}, \quad 0 \leq R \leq 1 \quad (2.60)$$

**Información en señales continuas.** Por cuanto la cantidad de valores, que se contienen en cualquier rango continuo, es infinita, así sea dado sobre un intervalo finito, entonces la cantidad de información en una variable continua también es infinita, así como su entropía. Por lo tanto, se define la *entropía diferencial* de la variable aleatoria continua  $x$ :

$$h(x) = E \left\{ \log \frac{1}{p(x)} \right\} = \int_X p(x) \log \frac{1}{p(x)} dx \quad (2.61)$$

la cual no se puede analizar como una medida de información propia, a diferencia de lo que ocurre con la entropía de las variables discretas. La medida  $h(x)$  no tiene las mismas propiedades que la entropía de las variables discretas, en particular la entropía diferencial puede tomar valores negativos. El sentido informativo no está en la misma entropía diferencial  $h(x)$ , sino en la diferencia de dos valores de  $h(x)$ , motivo por el cual se da el nombre de entropía diferencial.

**Ejemplo 2.19.** Hallar la entropía diferencial para los casos de FDP uniforme (2.33) y Gaussiana (2.23).

En correspondencia con (2.61), se tiene

– *FDP uniforme.*

$$h_U(x) = - \int_a^b p(x) \ln p(x) dx = - \frac{1}{b-a} \ln \frac{1}{b-a} \int_a^b dx = \ln(b-a)$$

Por cuanto se puede demostrar que  $b-a = 2\sqrt{3}\sigma_x$ , entonces,  $h_U(x) = \ln(2\sqrt{3}a)$

– *FDP Gaussiana.*

$$\begin{aligned} h_N(x) &= - \int_a^b \mathcal{U}(m_{1x}, \sigma_x) \ln \mathcal{U}(m_{1x}, \sigma_x) dx = \\ &= \ln(\sqrt{2\pi e}\sigma) \end{aligned}$$

Como se observa de las relaciones obtenidas, la diferencia de valores entre las ambas

entropías corresponde a

$$\Delta h = h_{\mathcal{N}}(x) - h_{\mathcal{U}}(x) = \ln \sqrt{2\pi e} \sigma_{\mathcal{N}(x)} - \ln (2\sqrt{3} \sigma_{\mathcal{U}(x)})$$

Si se asumen las varianzas de ambas distribuciones iguales  $\sigma_{\mathcal{U}(x)}^2 = \sigma_{\mathcal{U}(x)}^2 = \sigma_x^2$ , entonces el valor final de diferencia se obtiene como

$$\Delta h = \ln \frac{\sqrt{2\pi e}}{2\sqrt{3}} = 0.17$$

De los resultados obtenidos se observa que la entropía diferencial de una variable aleatoria continua, con función densidad de probabilidad Gaussiana o uniforme, no depende del valor medio de la variable aleatoria (por cuanto, un valor constante no lleva información) y solo depende en forma proporcional de la varianza de los valores aleatorios.

---

La propiedad de aditividad se conserva en el caso de la entropía diferencial. La entropía diferencial condicional de las variables  $\mathbf{x}$  e  $\mathbf{y}$ , definida como

$$h(\mathbf{x}|\mathbf{y}) = E \left\{ \log \frac{1}{p(\mathbf{x}|\mathbf{y})} \right\}$$

tiene las mismas propiedades que la entropía condicional de las variables discretas.

Así mismo, de todas las posibles FDP de la variable aleatoria  $\mathbf{x}$ , que tengan un valor fijo de potencia  $\sigma_x^2$ , la mayor entropía diferencial posible ocurre para la densidad de probabilidad Gaussiana.

## Problemas

**Problema 2.1.** Hallar el exceso  $\gamma_{2\xi}$  para las FDP: uniforme, Poisson, Binomial y Gaussiana.

**Problema 2.2.** Sea la variable aleatoria  $\eta$  una función lineal de  $\xi$ :  $\eta = \alpha\xi + \beta$ , donde  $\alpha$  y  $\beta$  son constantes. Encontrar la FDP de la variable  $\eta$  para el caso en que la densidad  $p(\xi)$  sea Gaussiana.

**Problema 2.3.** La variable  $\xi$  con FDP uniforme,  $p(\xi) = 1/(\alpha + \beta)$ ,  $\alpha < \xi < \beta$ ,  $\alpha < \beta$ , pasa por un detector cuadrático  $\eta = \xi^2$ . Determinar y representar:

- (a). La función de distribución  $F(\xi)$ .
- (b). La función de distribución y la función densidad de probabilidad de  $\eta$ .

**Problema 2.4.** La variable aleatoria  $\xi$  se describe por la FDP binomial. Hallar los valores de la media y la varianza de la magnitud  $\eta = \exp(\alpha\xi)$ , para los casos:  $p \neq q$  y  $p = q$ .

**Problema 2.5.** Sea la variable aleatoria  $\xi$  con FDP  $p(\xi) = \alpha \exp(-\beta|x|)$ ,  $-\infty < x < \infty$ , donde  $\alpha$  y  $\beta$  son constantes. Encontrar la relación que deben cumplir las constantes.

- (a). Calcular la función de distribución de la magnitud  $\xi$ .
- (b). Dibujar la función de distribución y la FDP para  $\beta = 2$ .

**Problema 2.6.** Sea la fase aleatoria  $\Phi$  uniformemente distribuida en el intervalo  $(0, 2\pi)$ . Calcular la FDP de la variable aleatoria de amplitud  $\xi(t) = A \cos^n(\omega t + \Phi)$ , donde  $\omega$  y  $t$  son constantes. Considérese  $n = 1$ ,  $n = 2$ .

**Problema 2.7.** La variable aleatoria  $\xi$  tiene FDP de Laplace,  $p(\xi) = 0.5t \exp(-t|\xi|)$ ,  $t > 0$ . Hallar la media y la varianza de la variable aleatoria.

**Problema 2.8.** Hallar  $\xi$  y  $\sigma_\xi^2$  de la variable entera aleatoria distribuida entre  $[0, N]$ , para la cual  $\sum_k^N p(\xi_k) = 1$ .

**Problema 2.9.** Hallar los momentos iniciales de la FDP de Weibull, definida en (2.29). Analizar los casos particulares de  $c = 1$  (FDP exponencial).

**Problema 2.10.** Demostrar que para la FDP de Cauchy no existe ningún momento inicial.

**Problema 2.11.** Sea  $\xi$  una variable aleatoria, para la cual converge el segundo momento inicial, asumiendo que  $c$  es una constante. Hallar para qué valor de  $c$  se obtiene el mínimo valor de  $m_{1\eta}$ , donde  $\eta = \xi - c$ .



## 2.2. Estimación en variables aleatorias

Hasta ahora, las definiciones, dadas para los diferentes valores y momentos de aleatoriedad, han supuesto que la cantidad de información, dispuesta sobre cualquier señal aleatoria, es completa. Debido a restricciones de costos, tiempo y posibilidad de tomar toda la información sobre los objetos pertenecientes a la señal aleatoria en análisis, en la realidad se tiene alguna trayectoria  $\mathbf{x}$  o conjunto de segmentos  $\{x_i\}$  tomados de la variable observada  $\mathbf{x} \in \xi$ . Por esto, es necesario calcular de manera aproximada cada valor de aleatoriedad a partir de la trayectoria disponible. En particular, a partir del análisis mediante una función  $g_\theta(x_i)$  sobre los resultados experimentales de una observación con longitud  $n$ ,  $\{x_i : i = 1, \dots, n\}$  se obtiene la característica estadística, o *estimación*,  $\tilde{\theta}$  del valor aleatorio  $\theta$ , entonces:

$$\tilde{\theta} = g_\theta(x_1, \dots, x_n) \quad (2.62)$$

Es necesario considerar, que al repetir el proceso de estimación, debido a la aleatoriedad que presentan los valores en cada  $k$  observación  $\{x_{lk} : l = 1, \dots, n\}$ , entonces la estimación  $\tilde{\theta} = g_\theta(x_{1k}, \dots, x_{nk})$  también es otra variable aleatoria.

Las características probabilísticas de  $\tilde{\theta}$  dependen de factores tales como, la estructura de aleatoriedad de cada variable  $\{x \in \xi : x_1, \dots, x_n\}$ , de la longitud de la observación  $n$  y del tipo de función  $g_\theta(x_i)$  empleada para el análisis de los datos. En este sentido, usualmente, la precisión en la aproximación del cálculo de la característica se describe por la *potencia del error* definida en la forma:

$$\varepsilon^2 = E\{(\tilde{\theta} - \theta)^2\} \quad (2.63)$$

que también corresponde a una variable aleatoria, motivo por el cual, a la función  $g_\theta(x_i)$  se le imponen las siguientes restricciones propias de todas las estimaciones:

1. *Ausencia de sesgo.* Caracteriza el comportamiento de la estimación cuando se amplía la longitud de la observación,  $n \rightarrow \infty$ :

$$E\{\tilde{\theta}\} = \lim_{N \rightarrow \infty} E\{g_\theta(x_1, \dots, x_n)\} = \theta$$

2. *Consistencia.* Analiza la convergencia en el sentido probabilístico del valor de la estimación al aumentar la longitud de la observación:

$$\lim_{n \rightarrow \infty} P\{|\tilde{\theta} - \theta| < \varepsilon\} = P\{|g_\theta(x_1, \dots, x_n) - \theta| < \varepsilon\} = 1$$

Basados en la desigualdad de Chebyshev (2.12), se puede demostrar que la condición suficiente para la consistencia de la estimación es [15]:

$$\lim_{n \rightarrow \infty} E\{(\tilde{\theta} - \theta)^2\} = 0$$

3. *Efectividad.* Cuando se tienen longitudes limitadas de la observación, los diversos

métodos de estimación conocidos pueden presentar diferentes valores de potencia de error (2.63). Entonces, se considera que entre menor sea el valor obtenido del error, para el correspondiente método de estimación, los valores calculados estarán agrupados en distancias más cercanas alrededor del valor  $\theta$ :

$$\epsilon^2 = \min_{\tilde{\theta}} E \{ (\tilde{\theta} - \theta)^2 \} \quad (2.64)$$

4. *Suficiencia.* Cuando la función densidad de probabilidad de la observación  $p(\theta; \tilde{\theta})$  se puede representar en la forma:

$$p(\theta; x_1, \dots, x_n) = p(\theta; \tilde{\theta}) h(x_1, \dots, x_n) = p(\theta; g_\theta(x_1, \dots, x_n)) h(x_1, \dots, x_n)$$

donde  $h(x_1, \dots, x_n)$  no depende de  $\theta$ . Esto es, toda la información contenida en la observación con relación al parámetro  $\theta$ , también está contenida en  $\tilde{\theta}$ .

En la práctica es usual el empleo del error relativo de estimación:

$$\Delta \tilde{\theta} = |\tilde{\theta} - \theta| \quad (2.65)$$

con lo que las primeras tres propiedades de la estimación se expresan en la respectiva forma:

- (a).  $E\{\Delta \tilde{\theta}\} = 0$
- (b).  $\lim_{n \rightarrow \infty} \sigma^2(\Delta \tilde{\theta}) = 0$
- (c).  $\sigma^2(\Delta \tilde{\theta}_k) = \min_{\tilde{\theta} \forall k} \sigma^2(\Delta \tilde{\theta})$

Debido a la naturaleza aleatoria de la estimación y, en particular de su potencia de error (2.63), es necesario determinar un intervalo de valores (y no solamente un valor), en el cual con un nivel de confianza, a priori dado, se localice el valor de la estimación. De otra parte, el intervalo con longitud  $(\theta_{\min}, \theta_{\max})$ , cuyo centro sirve de estimación de una característica aleatoria, el cual con una probabilidad de  $\alpha$  está contenido el valor verdadero de la característica, se denomina *intervalo de confianza* con valor de *significación*  $\alpha$ . En concordancia con lo anterior, existen dos clases de estimaciones: las *puntuales*, cuando el resultado de la aproximación corresponde a un valor y las estimaciones *de intervalo* cuando la aproximación se da dentro de un rango de valores.

### 2.2.1. Estimación puntual de momentos

La suposición, aunque sea con suficientes argumentos, sobre un tipo concreto de FDP,  $p_\xi(x)$ , aún no significa que se tengan los valores concretos de sus respectivos momentos, con lo cual, hay necesidad de estimarlos sobre los valores de la observación  $x_1, \dots, x_n$ .

Existen diversos métodos para la construcción de estimadores puntuales, sobre una observación dada, entre los principales están los siguientes:

**Métodos de los momentos.** Se basa en la equivalencia asumida de los momentos de estimación con los momentos de la FDP teórica, obteniéndose el sistema de ecuaciones:

$$m_{k\theta} \equiv \tilde{m}_{k\xi}; \quad k = 1, 2, \dots, d$$

en función de la cantidad  $d$  de las variables desconocidas,  $\theta_1, \theta_2, \dots, \theta_d$ .

En caso de existir una solución única,  $\tilde{\theta} = g_\theta(x_{1k}, \dots, x_{nk})$ , para el sistema de ecuaciones y cuando la función  $g_\theta$  es continua, entonces la estimación obtenida  $\tilde{\theta}_k$  es consistente. En particular, se puede tomar la siguiente equivalencia de momentos:

$$m_{k\theta} = \int_{x \in \xi} \xi^k p_\theta(d\xi) \cong \frac{1}{n} \sum_{l=1}^n x_l^k, \quad k = 1, 2, \dots, d \quad (2.66)$$

---

**Ejemplo 2.20.** Dada la observación  $\{x_k : 1, \dots, n\}$ , analizar las propiedades de estimación del método de momentos (2.66) para el caso de la media y la varianza.

La esperanza matemática del error (2.65) de estimación para un momento inicial de  $k$  orden se determina como:

$$\begin{aligned} \Delta \tilde{m}_{k\xi} &= \tilde{m}_{k\xi} - m_{k\theta} \\ E\{\Delta \tilde{m}_{k\xi}\} &= E\{\tilde{m}_{k\xi} - m_{k\theta}\} = E\left\{\frac{1}{n} \sum_{l=1}^n x_l^k\right\} - m_{k\theta} = \frac{1}{n} \sum_{l=1}^n E\{x_l^k\} - m_{k\theta} \\ &= \frac{1}{n} n m_{k\theta} - m_{k\theta} = 0 \end{aligned}$$

De lo anterior se deduce que la estimación  $\tilde{m}_{k\theta}$  no tiene sesgo.

De manera similar, se halla la varianza del error (2.65) de estimación:

$$\begin{aligned} \sigma^2(\Delta \tilde{m}_{k\xi}) &= \sigma^2(\tilde{m}_{k\xi}) = \sigma^2\left(\frac{1}{n} \sum_{l=1}^n x_l^k\right) = \frac{1}{n^2} n \sigma^2(x^k) = \frac{1}{n} E\{(x^k - m_{k\theta})^2\} \\ &= \frac{1}{n} E\{x^{2k} - 2m_{k\theta} x^k + m_{k\theta}^2\} = \frac{1}{n} E\{x^{2k} - m_{k\theta}^2\} = \frac{1}{n} (m_{2k\xi} - m_{k\theta}^2) \end{aligned}$$

Al hacer  $n \rightarrow \infty$  se tiene  $\sigma^2(\Delta \tilde{m}_{k\xi}) = 0$ . Por lo tanto, la estimación es consistente.

Haciendo  $k = 1$  se obtiene la estimación de la media, para la cual se tiene que

$$E\{\tilde{m}_{1\xi}\} = m_{1\xi}, \quad \sigma^2(\tilde{m}_{1\xi}) = \frac{m_{2\xi} - m_{1\xi}^2}{n} = \frac{\sigma_\xi^2}{n}$$

La estimación de los momentos centralizados se obtiene como:

$$\tilde{\mu}_{k\xi} = \frac{1}{n} \sum_{l=1}^n (x_l - \tilde{m}_{1\xi})^k$$

con error absoluto de estimación:  $\Delta \tilde{\mu}_{k\xi} = \tilde{\mu}_{k\xi} - \mu_{k\theta}$ . Se puede demostrar que para cualquier  $k$ ,  $E\{\Delta \tilde{\mu}_{k\xi}\} \neq 0$ . Sin embargo, de acuerdo a la última expresión obtenida para la varianza de estimación, se tienen los valores asintóticos:  $E\{\Delta \tilde{\mu}_{k\xi}\} \rightarrow 0, \sigma^2(\Delta \tilde{\mu}_{k\xi}) \rightarrow 0, \forall n \rightarrow \infty$ .

---

En general, las estimaciones obtenidas mediante el método de los momentos, (2.66), no son efectivas [17].

**Método de máxima verosimilitud.** Sea  $x \in \xi(\theta)$  una variable aleatoria con FDP que depende únicamente del parámetro  $\theta$ . Cuando se tiene la observación compuesta por los valores aleatorios independientes  $x_1, \dots, x_n$ , la probabilidad, de que cualquier posible trayectoria conste precisamente de estos  $n$  valores discretos, está dada por el producto:

$$\lambda(\theta) = p(x_1, \theta) p(x_2, \theta) \cdots p(x_n, \theta)$$

Mientras, en el caso de variables continuas, la probabilidad de que la observación conste de  $n$  valores en intervalos pequeños  $x_1 \leq x \leq x_1 + \Delta x, \dots, x_n \leq x \leq x_n + \Delta x$  está dada por la expresión:

$$\lambda(\Delta x, \theta)^n = p(x_1, \theta) \Delta x \cdots p(x_n, \theta) \Delta x = \Delta x^n \prod_i^n p(x_i, \theta) \quad (2.67)$$

El valor de  $\lambda$  se denomina *función de verosimilitud* que depende de cada trayectoria  $x_1, \dots, x_n$  y del parámetro desconocido  $\theta$ . En el método de verosimilitud se escoge la aproximación del valor desconocido de  $\theta$ , para un valor de  $\lambda$  tan grande como sea posible. Si  $\lambda$  es una función derivable de  $\theta$ , entonces, una condición necesaria para que  $\lambda$  tenga un máximo está en que

$$d\lambda/d\theta = 0$$

Cualquier solución de la anterior ecuación corresponde a la *estimación de máxima verosimilitud* para el parámetro  $\theta$ , en la cual se reemplazan los valores de la observación por cada una de las variables independientes aleatorias  $\xi_1, \dots, \xi_n$  para obtener la variable aleatoria  $\tilde{\theta}$  denominada *estimador de máxima verosimilitud* con respecto a  $\theta$ . Cuando la FDP incluye  $r$  parámetros desconocidos  $\theta_1, \dots, \theta_r$ , es preferible conformar un sistema de  $r$  ecuaciones en la forma:

$$\frac{d \ln(\lambda)}{d\theta_1} = 0, \dots, \frac{d \ln(\lambda)}{d\theta_r} = 0, \quad (2.68)$$

La ventaja de esta transformación está que las derivadas de los productos se reemplazan por la derivación de sumas. No obstante, en algunos casos, la solución del sistema es difícil de lograr.

Las estimaciones de máxima verosimilitud tienen las siguientes propiedades:

- Cuando existe la estimación suficiente  $\tilde{\theta}(\xi_1, \xi_2, \dots, \xi_N)$  para el parámetro  $\theta$ , entonces, cada solución de la ecuación de verosimilitud es función de  $\tilde{\theta}(\xi_1, \xi_2, \dots, \xi_N)$ .
- Si para  $\theta$  existe una estimación efectiva  $\tilde{\theta}(\xi_1, \xi_2, \dots, \xi_N)$ , entonces, la ecuación de verosimilitud, (2.68) tiene una única solución:  $\tilde{\theta}(\xi_1, \xi_2, \dots, \xi_n)$ .

**Teorema 2.6.** Sea la FDP  $p(\theta, \xi)$ , tal que cumpla las siguientes condiciones:

1. Para casi todos los  $\xi$  existen las derivadas:

$$\frac{\partial^k \ln p(\theta, \xi)}{\partial \theta^k}, \quad k = 1, 2, 3;$$

2. Se cumple la desigualdad:

$$\left| \frac{\partial^k \ln p(\theta, \xi)}{\partial \theta^k} \right| \leq G_k(\xi), \quad k = 1, 2, 3;$$

siendo las funciones  $G_k(\xi)$ ,  $k = 1, 2$  integrables, esto es,

$$\sup_{\theta} \int_{\Xi} G_k(\xi) p(\theta, \xi) d\xi < \infty$$

3. Para cada  $\theta$  la siguiente integral

$$I(\theta) = \int_{\Xi} \left( \frac{\partial \ln p(\theta, \xi)}{\partial \theta} \right)^2 p(\theta, \xi) d\xi \quad (2.69)$$

es positiva y acotada. Entonces, la ecuación de verosimilitud (2.68) tiene solución  $\tilde{\theta}(\xi_1, \dots, \xi_N)$  que es una estimación consistente y asintóticamente efectiva.

**Ejemplo 2.21.** Sea la trayectoria de valores aleatorios estadísticamente independientes,  $\{10, 11, 8, 12, 9\}$ , de un proceso que se asume con FDP exponencial,  $p(\theta, \xi) = \frac{1}{\theta} \exp(-\xi/\theta) u(\xi)$ ,  $\theta > 0$ . Hallar la estimación de máxima verosimilitud para  $\theta$ .

La función de verosimilitud de  $\theta$  es igual a

$$\begin{aligned} \lambda(\theta) &= \prod_{n=1}^N p(\theta, \xi_n) = \prod_{n=1}^5 \frac{1}{\theta} \exp\left(-\frac{\xi_n}{\theta}\right) \\ &= \theta^{-5} \exp\left(-\frac{50}{\theta}\right), \quad \theta > 0 \end{aligned}$$

Al derivar la anterior expresión por el parámetro  $\theta$  se obtiene:

$$\frac{d\lambda}{d\theta} = -5\theta^{-6} \exp\left(-\frac{50}{\theta}\right) + \theta^{-5} \left(\frac{50}{\theta^2}\right) \exp\left(-\frac{50}{\theta}\right)$$

Cuando se hace la derivada igual a 0, al resolver por  $\tilde{\theta}$ , se obtiene el valor concreto de la estimación:  $\tilde{\theta} = 50/5 = 10$ .

En el caso de tener una variable aleatoria  $\xi$  con FDP normal, con valores desconocidos de media y varianza, la función de verosimilitud se determina por la expresión:

$$\lambda = \frac{1}{\sqrt{2\pi}\sigma_{\xi}} \cdots \frac{1}{\sqrt{2\pi}\sigma_{\xi}} e^{-\frac{(\xi_1 - m_{1\xi})^2}{2\sigma_{\xi}^2} - \cdots - \frac{(\xi_N - m_{1\xi})^2}{2\sigma_{\xi}^2}} = \left( \frac{1}{\sqrt{2\pi}\sigma_{\xi}} \right)^{N/2} e^{-h}, \quad (2.70)$$

donde  $h = -\frac{1}{2\sigma_\xi^2} \sum_{k=1}^N (\xi_k - m_{1\xi})^2$ .

Hallando el logaritmo se tiene:

$$A_\xi = \ln(\lambda) = -\frac{N}{2} \ln(\sqrt{2\pi}) - \frac{N}{2} \ln(\sigma_\xi^2) - h \quad (2.71)$$

El máximo del funcional  $\ln(\lambda)$  por el parámetro  $m_{1\xi}$  corresponde a hallar el mínimo del último término (dependiente del parámetro desconocido  $m_{1\xi}$ ), que contiene la suma de los cuadrados  $\sum_{k=1}^N (\xi_k - m_{1\xi})^2$ . Por lo tanto, en el caso concreto de la FDP normal, el método de máxima verosimilitud coincide con el método de los mínimos cuadrados, mostrado en el ejemplo 2.23. Así, para hallar la estimación de la media, se deriva e iguala a cero la ecuación (2.71) por el parámetro desconocido  $m_{1\xi}$ ,

$$\frac{d \ln(\lambda)}{d m_{1\xi}} = \frac{1}{\sigma_\xi^2} \sum_{k=1}^N (\xi_k - m_{1\xi}) = 0,$$

entonces,

$$\sum_{k=1}^N (\xi_k - m_{1\xi}) = \sum_{k=1}^N \xi_k - N m_{1\xi} = 0$$

Al resolver para  $m_{1\xi}$ , expresando la solución en términos de  $\tilde{m}_{1\xi}$ , se obtiene,

$$\tilde{m}_{1\xi} = \frac{1}{N} \sum_{k=1}^N \xi_k \quad (2.72)$$

De igual manera, se obtiene la ecuación para la estimación de la varianza,

$$\frac{d \ln(\lambda)}{d \sigma_\xi^2} = -\frac{1}{2\tilde{\sigma}_\xi^2} \left( N - \frac{1}{\tilde{\sigma}_\xi^2} \sum_{k=1}^N (\xi_k - \tilde{m}_{1\xi})^2 \right) = 0 \quad (2.73)$$

La solución para  $\sigma_\xi^2$ , en términos de  $\tilde{\sigma}_\xi^2$ , resulta en la respectiva estimación:

$$\tilde{\sigma}_\xi^2 = \frac{1}{N} \sum_{k=1}^N (\xi_k - \tilde{m}_{1\xi})^2$$

**Método del mínimo  $\chi^2$ .** Sean  $\xi_1, \xi_2, \dots, \xi_N$  las observaciones independientes correspondientes a la variable aleatoria  $\xi$ , que contienen el parámetro desconocido  $\theta$  y están dadas sobre un espacio  $X$ , tal que pueda ser dividido en  $r$  conjuntos incongruentes de la forma,  $X_1, X_2, \dots, X_r$ . Se asume que la cantidad de observaciones de la trayectoria  $\xi_1, \xi_2, \dots, \xi_N$ , que se localizan dentro del conjunto  $\xi_i$  corresponde a  $N_i$ . Si el conjunto  $X$  es finito (o sea, la variable aleatoria puede tomar una cantidad finita de valores), entonces, se puede considerar que cada  $X_i$  es un conjunto conformado por un solo punto.

De esta manera, se realiza la agrupación de los resultados de la observación, formando la siguiente variable aleatoria (2.31),

$$\chi^2 = \sum_{i=1}^N \frac{(N_i - Np_i(\theta))^2}{Np_i(\theta)}$$

donde  $p_i(\theta) = p_\theta(X_i)$  ( $i = 1, 2, \dots, r$ ). En este caso, la estimación  $\tilde{\theta}(\xi_1, \xi_2, \dots, \xi_N)$  se denomina *estimación por el método del mínimo  $\chi^2$* , la cual se obtiene al minimizar por  $\theta$  la magnitud  $\chi^2$ . Asumiendo que  $\theta$  es un parámetro con dimensión  $d$ , entonces, para encontrar la estimación por el método del mínimo  $\chi^2$  se conforma el sistema de ecuaciones:

$$\sum_{i=1}^N \left( \frac{N_i - Np_i(\theta)}{p_i(\theta)} + \frac{(N_i - Np_i(\theta))^2}{2Np_i(\theta)} \right) \frac{\partial p_i(\theta)}{\partial \theta^k} = 0, \quad k = 1, 2, \dots, d$$

Cabe decir que las propiedades asintóticas de las estimaciones del método del mínimo  $\chi^2$  son muy cercanas a las estimaciones de máxima verosimilitud.

### 2.2.2. Intervalos de confianza

En algunos casos, es importante además de la estimación del parámetro desconocido, determinar la región donde se presume se encuentra el verdadero valor del parámetro. Aunque esa región se construye sobre la observación, sus valores cambian de una a otra trayectoria en forma aleatoria. Por lo tanto, se puede definir la probabilidad con que en una región concreta se localice el verdadero valor del parámetro. Escogiendo un cierto valor suficientemente pequeño  $\alpha > 0$ , que corresponde al valor de significación, se puede construir una regla que permita para cada trayectoria estimar la región de medida, en la cual con una probabilidad de  $1 - \alpha$  el valor del parámetro se encuentra. Esta región se denomina *intervalo de confianza*, mientras el valor  $1 - \alpha$  es el *coeficiente o nivel de confianza*.

Sean  $x_1, x_2, \dots, x_n$  las observaciones independientes de la variable aleatoria  $\xi$ , cuya densidad contiene el parámetro desconocido  $\theta \in \mathbb{R}$ , para el cual se tiene la estimación  $\tilde{\theta}(x_1, x_2, \dots, x_n)$ . Sea  $q_\theta(dt)$  una densidad de la estimación  $\tilde{\theta}$ , suponiendo que el valor verdadero del parámetro coincide con  $\tilde{\theta}$ , esto es,

$$q_\theta(dt) = Q_\theta \left\{ \tilde{\theta}(x_1, x_2, \dots, x_n) \in dt \right\}$$

donde  $Q_\theta(dx_1, dx_2, \dots, dx_n) \in \mathbb{R}^n$  es la distribución, que se determina por:

$$Q_\theta(dx_1, dx_2, \dots, dx_n) = P_\theta(dx_1) P_\theta(dx_2) \cdots P_\theta(dx_n) \quad (2.74)$$

Sea  $q_\theta(dt)$  una densidad sin átomos, entonces para un valor dado de  $\alpha$  se pueden escoger los valores  $a_1(\theta, \alpha)$ ,  $a_2(\theta, \alpha)$ , donde  $a_1(\theta, \alpha) < a_2(\theta, \alpha)$ , para los cuales se cumple,

$$\int_{\{t < a_1(\theta, \alpha)\}} q_\theta(dt) + \int_{\{t > a_2(\theta, \alpha)\}} q_\theta(dt) = \alpha$$

Este método de selección no tiene solución única. De otra parte, asumiendo la continuidad por  $\theta$  de las funciones a escoger y que cada una de las ecuaciones  $a_i(\theta, \alpha) = \lambda$  ( $i = 1, 2$ ) tenga solución única  $c_i(\lambda, \alpha)$  ( $i = 1, 2$ ), entonces, las relaciones son equivalentes:

$$\begin{aligned} Q_\theta \{a_1(\theta, \alpha) < \tilde{\theta} < a_2(\theta, \alpha)\} &= 1 - \alpha \\ Q_\theta \{c_1(\tilde{\theta}, \alpha) < \theta < c_2(\tilde{\theta}, \alpha)\} &= 1 - \alpha \end{aligned}$$

De esta manera, conociendo la distribución de la estimación  $\tilde{\theta}(x_1, x_2, \dots, x_n)$ , para un valor dado de  $\alpha > 0$  se puede construir el intervalo de confianza entre los valores  $(c_1(\tilde{\theta}, \alpha), c_2(\tilde{\theta}, \alpha))$  para un nivel de confianza  $1 - \alpha$ .

Si  $q_\theta(dt)$  tiene átomos, los valores  $a_1(\theta, \alpha)$  y  $a_2(\theta, \alpha)$  se escogen de la desigualdad:

$$\int_{\{t < a_1(\theta, \alpha)\}} q_\theta(dt) + \int_{\{t < a_2(\theta, \alpha)\}} q_\theta(dt) \leq \alpha$$

La desigualdad está dada, por cuanto, no pueden existir valores de  $a_1$  y  $a_2$ , para los cuales tenga lugar la igualdad [17]. Seguidamente, el proceso es similar al descrito anteriormente para la obtención del intervalo de confianza. En forma general, al tomar como base diversas estimaciones  $\tilde{\theta}$  pueden darse diferentes intervalos de confianza, pero en cualquier caso, es importante que su longitud sea lo más pequeña posible. Por esto, en su construcción se emplean estimaciones efectivas o, por lo menos, asintóticamente efectivas que sean obtenidas, por ejemplo del método de máxima verosimilitud.

La construcción del intervalo de confianza asume una distribución  $P_\theta(dx)$  sin átomos y que se puede hallar una función  $g(\theta, x_1, x_2, \dots, x_n)$ ,  $x_i \in \xi$ , con las siguientes propiedades:

- La función  $g(\theta, x_1, x_2, \dots, x_n)$ , es continua y monótona con relación a  $\theta$ .
- La función de distribución  $Q_\theta\{g(\theta, x_1, x_2, \dots, x_n) < \gamma\}$  no depende de  $\theta$ ,  $\forall \gamma \in \mathbb{R}$ .

Para un valor dado de  $\alpha > 0$ , se escogen los valores  $a_1(\alpha)$  y  $a_2(\alpha)$ , tales que:

$$Q_\theta\{a_1(\alpha) < g(\theta, x_1, x_2, \dots, x_n) < a_2(\alpha)\} = 1 - \alpha$$

De acuerdo con la propiedad b, los valores  $a_i(\alpha)$  ( $i = 1, 2$ ) no dependen de  $\theta$ . Al notar por  $c_i$  ( $i = 1, 2$ ) los valores que satisfacen las relaciones  $g(\theta, x_1, x_2, \dots, x_n) = a_i(\alpha)$ , respectivas, que dependen de la observación  $\theta, x_1, x_2, \dots, x_n$  y  $\alpha$ , se obtiene que

$$Q_\theta(c_1 < \theta < c_2) = 1 - \alpha$$

por lo tanto,  $(c_1, c_2)$  corresponde al intervalo de confianza con nivel de confianza  $1 - \alpha$ .

Así por ejemplo, sea dada la distribución integral  $F_\theta(x)$  continua y monótona para  $\theta$ ,

$$F_\theta(x) = \int_{-\infty}^x p_\theta(d\xi), \quad x \in \mathbb{R}$$



Entonces, se puede comprobar que la función  $\prod_{k=1}^n F_\theta(x_k)$  cumple las condiciones  $a$  y  $b$ , por lo que esta se puede emplear para la construcción de los intervalos de confianza. Debido a la continuidad de  $F_\theta(\xi)$ :

$$Q_\theta\{F_\theta(x_k) < \gamma\} = \begin{cases} 0, & \gamma \leq 0 \\ \gamma, & 0 \leq \gamma \leq 1 \\ 1, & \gamma > 1 \end{cases}$$

donde la variable correspondiente a la suma  $\sum_{k=1}^n \log F_\theta(x_k)$  tiene FDP con  $\Gamma$ -densidad:

$$Q_\theta\left\{-\log a_2 < -\sum_{k=1}^n \log F_\theta(x_k) < -\log a_1\right\} = \int_{-\log a_2}^{-\log a_1} \frac{1}{\Gamma(n)} \lambda^{n-1} e^{-\lambda} d\lambda$$

Se pueden escoger los valores  $a_1$  y  $a_2$  ( $a_1 < a_2$ ) para un valor dado  $\alpha > 0$ , de tal manera que la integral de la derecha en la última igualdad sea igual a  $1 - \alpha$ , obteniéndose que:

$$Q_\theta\left\{a_1 < \prod_{k=1}^n F_\theta(x_k) < a_2\right\} = 1 - \alpha$$

Por cuanto, la función  $\prod_{k=1}^n F_\theta(x_k)$  es monótona por  $\theta$ , entonces, existen ciertos valores  $c_1$  y  $c_2$ , que dependen solamente de  $\alpha$  y  $x_1, \dots, x_n$ , tales que determinen el intervalo de confianza  $(c_1, c_2)$  para  $\theta$  con nivel de confianza  $1 - \alpha$ .

En la práctica, es frecuente el uso de los *percentiles* de nivel  $1 - \alpha$ , entendidos como el valor  $x_{1-\alpha}$  para el cual  $F_\theta(x_{1-\alpha}) = 1 - \alpha$ . Así por ejemplo, en el caso de la FDP normal estándar,  $\mathcal{N}(0, 1)$ , el percentil de nivel  $\alpha = 1 - \varepsilon$  se calcula como:

$$\Phi(\xi_\alpha) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\xi_\alpha} e^{-\lambda^2/2} d\lambda = \alpha$$

**Estimación Bayesiana de intervalos.** Sea el parámetro aleatorio  $\theta$  con FDP a priori  $p(\theta)$ , cuya estimación  $\tilde{\theta}(x_1, \dots, x_n)$  está construida sobre la trayectoria de valores independientes  $x_1, \dots, x_n$ . Se asume que la distribución  $Q_\theta(dt)$  de la estimación  $\tilde{\theta}$  es absolutamente continua, como también lo es la respectiva función  $g(\theta, t)$  que en este caso se toma igual a la FDP condicional definida en (2.41), la cual de acuerdo con (2.4), se determina como,

$$g(\theta, \tilde{\theta}) = p(\theta|\tilde{\theta}) = \frac{p(\theta)p(\tilde{\theta}|\theta)}{\int_{\Theta} p(\theta)p(\tilde{\theta}|\theta)d\theta}$$

donde  $p(\tilde{\theta}|\theta)$  es la FDP condicional de la estimación  $\tilde{\theta}$  para un valor fijo de  $\theta$ , con lo cual la probabilidad de que el parámetro  $\theta$  esté localizado dentro de los límites  $(c_1, c_2)$  está dada

por la expresión,

$$P(c_1 \leq \theta \leq c_2 | \tilde{\theta}) = \int_{c_1}^{c_2} p(\theta | \tilde{\theta}) d\theta$$

Luego, dado el valor de  $\alpha$  se pueden determinar los valores límites,  $c_1(\tilde{\theta}, \alpha)$  y  $c_2(\tilde{\theta}, \alpha)$ , para los cuales se cumpla la igualdad:

$$P(c_1(\tilde{\theta}, \alpha) \leq \theta \leq c_2(\tilde{\theta}, \alpha) | \tilde{\theta}) = 1 - \alpha$$

### 2.2.3. Estimación de parámetros en la distribución normal

**Estimación de la media cuando se conoce la varianza.** Sea el conjunto de valores independientes  $x_1, \dots, x_n$ , de una trayectoria que se considera suficientemente grande, así que  $n > 30$ , correspondientes a la variable aleatoria normal  $x \in \xi$  con FDP, descrita como  $p(\theta, \xi) = \mathcal{N}(\theta, \sigma_\xi^2)$ , donde  $\theta$  es el parámetro desconocido, mientras  $\sigma$  conocida.

Sea la estimación de la esperanza matemática obtenida por el método de los momentos en (2.66),  $k = 1$ . De los resultados obtenidos en el ejemplo 2.20, se deduce que  $\tilde{\theta}$  es una variable aleatoria con distribución normal y momentos:

$$\begin{aligned} \theta(x_1, \dots, x_n) &= \tilde{m}_{1\xi} = m_{1\xi}, \\ \tilde{\sigma}_\xi^2 &= \sigma_\xi^2 / n \end{aligned}$$

luego, la estimación  $\tilde{\theta}$  no presenta sesgo, es consistente y, adicionalmente, efectiva [17]. En concordancia con lo anterior, la FDP es Gaussiana con parámetros  $p(\theta, \xi) = \mathcal{N}(m_{1\xi}, \sigma_\xi^2)$ . De esta manera, para un valor dado  $\alpha > 0$ , teniendo en cuenta (2.74) se deben hallar los valores ( $c_2 = -c_1 = c_\alpha$ ), tales que se cumpla la igualdad:

$$\begin{aligned} Q_\theta \{ \tilde{m}_{1\xi} - c_\alpha < \xi < \tilde{m}_{1\xi} + c_\alpha \} &= P \{ \tilde{m}_{1\xi} - c_\alpha < \xi < \tilde{m}_{1\xi} + c_\alpha \} = P \{ |\xi - \tilde{m}_{1\xi}| < c \} \\ &= \alpha \end{aligned}$$

Teniendo en cuenta la definición de FDP para la variable  $x \in \xi$  se obtiene:

$$\begin{aligned} Q_\theta(dx_1, \dots, dx_n) &= (2\pi\sigma_\xi^2)^{-\frac{n}{2}} \exp \left\{ -\frac{1}{2\sigma_\xi^2} \sum_{k=1}^n (x_k - \theta)^2 \right\} (dx_1, \dots, dx_n) \\ &= \int_{\tilde{m}_{1\xi} - c_\alpha}^{\tilde{m}_{1\xi} + c_\alpha} p_\xi(x) dx = 2\Phi \left( \frac{c_\alpha}{\sqrt{\sigma_\xi^2}} \right) - 1 = 2\Phi \left( c \sqrt{\frac{n}{\tilde{\sigma}_\xi^2}} \right) - 1 \\ &= \alpha \end{aligned}$$

siendo  $\Phi(x)$  la integral de Laplace, definida en (2.27). Así, la distancia obtenida, entre los puntos  $(\tilde{m}_{1\xi} - c, \tilde{m}_{1\xi} + c)$ , corresponde al intervalo de confianza con nivel  $1 - \alpha$ .

**Estimación de la varianza para una media conocida.** En este caso,

$$p(\theta, \xi) = \frac{1}{\sqrt{2\pi\theta}} \exp \left\{ -\frac{(\xi - m_{1\xi})^2}{2\theta} \right\}, \quad \xi \in \mathbb{R}$$

donde la varianza de la estimación  $\tilde{\theta}$ , de acuerdo con el ejemplo 2.20, se determina por la expresión  $E_{\theta}\{\tilde{\theta} - \theta\}^2 = \theta^2/n$ , entonces, se puede demostrar, que el valor  $n\tilde{\sigma}_{\xi}^2/\sigma_{\xi}^2$  tiene  $\chi^2$ -FDP con  $n$  grados de libertad.

El intervalo de confianza con nivel  $1 - \alpha$  es el cálculo de los valores  $a_1$  y  $a_2$ , tales que:

$$Q_{\theta} \left\{ a_1 < n\tilde{\sigma}_{\xi}^2/\sigma_{\xi}^2 < a_2 \right\} = 1 - \alpha$$

donde  $Q_{\theta}(dx_1, \dots, dx_n) = (2\pi\theta)^{-n/2} \exp \left\{ \frac{1}{2\theta} \sum_{k=1}^n (x_k - m_{\xi})^2 \right\} (dx_1, \dots, dx_n)$ . Luego,

$$\begin{aligned} Q_{\theta} \left( n \frac{\tilde{\theta}}{a_2} < \theta < n \frac{\tilde{\theta}}{a_1} \right) &= Q_{\theta} \left( \frac{n\tilde{\sigma}_{\xi}^2}{a_2} < \sigma_{\xi}^2 < \frac{n\tilde{\sigma}_{\xi}^2}{a_1} \right) \\ &= 1 - \alpha \end{aligned}$$

tal que  $(n\tilde{\sigma}_{\xi}^2/a_2, n\tilde{\sigma}_{\xi}^2/a_1)$  es el intervalo de confianza buscado.

**Estimación de la varianza cuando se conoce la media.** Sea la estimación  $\tilde{\theta} = \tilde{m}_{1\xi}$  sin sesgo y consistente. La construcción del intervalo de confianza parte del hecho de que la variable  $(\tilde{m}_{1\xi} - \theta)/\tilde{\mu}_{2\xi}$ , donde  $\tilde{\mu}_{2\xi}$  está dada en el ejemplo 2.20 ( $k = 2$ ), tiene FDP del tipo *Student* con  $n - 1$  grados de libertad, descrita como

$$p_{n-1}(x) = \frac{\Gamma\left(\frac{n}{2}\right)}{\sqrt{n}\Gamma\left(\frac{(n-1)}{2}\right)} (1+x^2)^{-\frac{n}{2}}$$

Con lo cual, el valor  $c_{\alpha}$  se determina de la relación

$$2 \int_0^{c_{\alpha}} p_{n-1}(x) dx = 1 - \alpha$$

Por lo que se obtiene  $Q_{\theta} \left\{ -c_{\alpha} < \frac{\tilde{m}_{1\xi} - \theta}{\sqrt{\tilde{\mu}_{2\xi}}} < c_{\alpha} \right\} = 1 - \alpha$ , donde

$$Q_{\theta}(dx_1, \dots, dx_n) = (2\pi\sigma_{\xi}^2)^{-n/2} \exp \left\{ -\frac{1}{2\sigma_{\xi}^2} \sum_{k=1}^n (x_k - \theta)^2 \right\} (dx_1, \dots, dx_n)$$

De esta manera, el intervalo de confianza para la estimación del parámetro  $\theta$  con nivel  $1 - \alpha$  corresponde al segmento  $(\tilde{m}_{1\xi} - \sqrt{\tilde{\mu}_{2\xi}}c_{\alpha}, \tilde{m}_{1\xi} + \sqrt{\tilde{\mu}_{2\xi}}c_{\alpha})$ .

**Estimación conjunta de la varianza y la media.** Sean desconocidos los parámetros de la FDP normal,  $m_{1\xi}$  y  $\sigma_\xi^2$ , estimados, bien sea por el método de los momentos o el de máxima verosimilitud:

$$\tilde{m}_{1\xi} = \frac{1}{n} \sum_{k=1}^n x_k, \quad \tilde{\sigma}_\xi^2 = \tilde{\mu}_{2\xi} = \frac{1}{n} \sum_{k=1}^n (x_k - \tilde{m}_{1\xi})^2$$

En este caso, la estimación de  $\sigma_\xi^2$  resulta sesgada, por lo que hay necesidad de introducir el factor  $n/n - 1$ , para que la corrección del sesgo  $\tilde{\sigma}_\xi^2 = n\tilde{\sigma}_\xi^2/n - 1$ , entonces,  $(\tilde{m}_{1\xi}, \tilde{\sigma}_\xi^2)$  corresponden a las respectivas estimaciones no sesgadas de los parámetros  $(m_{1\xi}, \sigma_\xi^2)$ .

#### 2.2.4. Prueba de hipótesis

Sea una estructura estadística relacionada con la variable  $x \in \xi$ , para la cual la función de densidad de probabilidad  $p_\xi(x)$  es desconocida. Cualquier suposición que defina unívocamente a una clase de pertenencia la densidad  $p_\xi(x)$  se denomina *hipótesis* [18]. Una hipótesis  $H$  se determina como *simple* si el conjunto  $H$  consiste de un único parámetro  $\theta$ . En caso contrario,  $H$  es una hipótesis *compuesta*. En la práctica, se hace necesario la verificación en la correspondencia entre los resultados reales del experimento con la hipótesis que se tiene sobre su estructura aleatoria. En este sentido, se realiza el procedimiento de verificación o *prueba de hipótesis*, el cual permite de las observaciones analizadas tomar o rechazar una hipótesis dada.

La prueba estadística de hipótesis pertenece a la clase de estimación puntual y se realiza dividiendo el espacio de observaciones  $x \in \xi$  en dos espacios que no se intercepten:  $\mathfrak{X}_0 \cap \mathfrak{X}_1 = \emptyset$ , correspondientes a las dos hipótesis alternativas. Si los resultados de las observaciones pertenecen a  $x \in \mathfrak{X}_0$  (intervalo de toma de hipótesis), entonces se considera que la respectiva hipótesis se confirma por los datos empíricos y, por lo tanto, se acepta  $H_0$  como cierta. En caso contrario, cuando  $x \notin \mathfrak{X}_0$  (intervalo crítico), entonces, la hipótesis dada  $H_0$  no corresponde a las observaciones experimentales y, por lo tanto, se rechaza. Por cuanto, los valores de cada trayectoria de  $x \in \xi$  son aleatorios, se definen las respectivas probabilidades a priori de acierto de la prueba de la hipótesis  $P_\theta(x \in \mathfrak{X}_0)$  y  $P_\theta(x \notin \mathfrak{X}_0)$ , dado un valor de  $\theta$ . Sin embargo, la misma prueba de hipótesis puede generar errores de primer género (rechazar una hipótesis cuando realmente era cierta) y de segundo género (tomar la hipótesis como cierta cuando realmente era falsa), entonces, el procedimiento de prueba se debe optimizar para obtener el valor mínimo de ambos géneros de error.

Cabe anotar que, en la mayoría de los casos prácticos, es imposible optimizar el procedimiento de la verificación de la hipótesis, en el sentido de minimizar ambos géneros de error para cualquier valor tan pequeño como se quiera [17].

El valor máximo permitido  $\alpha$  del error de primer género para un criterio dado se denomina *valor de significación del criterio*,  $\sup_{\theta \in H} P_\theta(X_0) \leq \alpha$ . En el caso del error de segundo género, en vez de analizar su valor, notado por  $\beta$ , se analiza la magnitud,  $w = 1 - \beta$ , denominada *potencia del criterio* de la prueba de hipótesis  $H_0$  contra la hipótesis alternativa  $H_1$

La comparación de los diferentes criterios propuestos se realiza empleando métricas de efectividad asintóticas, basadas en el estudio de la convergencia de la función de potencia en las cercanías del parámetro  $\theta \in H$ . Se considera óptimo aquel criterio, que para un valor dado de significación  $\alpha$ , brinde el máximo valor de potencia del criterio, el cual define la siguiente función crítica  $\varphi(x)$ :

$$\int_{\xi} \varphi(x) p_0(x) dx = \alpha, \quad \varphi(x) = \begin{cases} 1, & p_1(x) \geq \lambda_{\alpha} p_0(x) \\ 0, & p_1(x) < \lambda_{\alpha} p_0(x) \end{cases} \quad (2.75)$$

siendo  $p_0(x)$  y  $p_1(x)$  las respectivas FDP de  $P_0$  y  $P_1$ .

En la práctica, para el desarrollo del criterio (2.75), denominado de *Neyman-Pearson*, en calidad de estadística suficiente se emplea la relación de verosimilitud (2.67), que implica la verificación de la desigualdad  $p_1(x) \geq \lambda_{\alpha} p_0(x)$  para una observación concreta  $x \in \xi$ . Si la desigualdad se cumple, entonces la hipótesis  $H_0$  se rechaza, en caso contrario se acepta. La constante  $\lambda_{\alpha} \in \mathbb{R}$  se determina a partir de la condición:  $P(\Lambda_x \geq \lambda_{\alpha}) = \alpha$ , siendo  $\Lambda_x = p_1(x)/p_0(x)$  la respectiva relación de verosimilitud.

**Ejemplo 2.22.** Dada la trayectoria de la variable aleatoria Gaussiana  $x \in \xi$  de longitud  $n$  y varianza conocida  $\sigma_{\xi}^2$ , realizar la prueba de hipótesis  $H_0$  sobre el valor medio de una variable aleatoria,  $m_{1\xi} = m_0$  contra la hipótesis alternativa  $H_1$  sobre el valor  $m_{1\xi} = m_1$  para el criterio de Neyman-Pearson empleando la estadística suficiente (2.67).

La función de verosimilitud del parámetro de distribución  $m_{1\xi}$  para ambas hipótesis, teniendo en cuenta (2.70), se determina como

$$\lambda(m_i; x) = (2\pi\sigma_{\xi}^2)^{-n/2} \exp\left(-\frac{1}{2\sigma_{\xi}^2} \sum_{k=1}^n (x_k - m_i)^2\right), \quad i = 0, 1$$

con lo cual la relación de verosimilitud es igual a

$$\Lambda_x = \frac{m_1 - m_0}{\sigma_{\xi}^2} \sum_{k=1}^n x_k - \frac{n(m_1^2 - m_0^2)}{2\sigma_{\xi}^2},$$

cuyo valor se compara con el umbral  $h_0 = \ln h$ , en concordancia con (2.71). La hipótesis nula se toma en favor de  $m_{1\xi} = m_0$  y se acepta como cierta si se cumple que,

$$\frac{m_1 - m_0}{\sigma_{\xi}^2} \sum_{k=1}^n x_k - \frac{n(m_1^2 - m_0^2)}{2\sigma_{\xi}^2} \underset{H_0}{\overset{H_1}{\gtrless}} \ln h \quad (1)$$

La desigualdad (1) se puede escribir en formas de las siguientes desigualdades, que dependen de la relación entre los valores  $m_0$  y  $m_1$ :

$$\begin{aligned} m_0 < m_1 : \quad \tilde{m}_{1\xi} &= \frac{1}{n} \sum_{k=1}^n x_k < \frac{m_1 - m_0}{2} + \frac{\sigma_{\xi}^2 \ln h}{n(m_1 - m_0)} = h_1 \\ m_0 > m_1 : \quad \tilde{m}_{1\xi} &= \frac{1}{n} \sum_{k=1}^n x_k > \frac{m_1 - m_0}{2} + \frac{\sigma_{\xi}^2 \ln h}{n(m_1 - m_0)} = h_2 \end{aligned}$$

De las anteriores desigualdades, se observa que para tomar la decisión sobre el valor medio es necesario comparar su valor estimado con los umbrales  $h_i, i = 1, 2$ , los cuales se escogen de acuerdo al criterio de optimización empleado.

### 2.2.5. Estimación de dependencias funcionales

Sea  $\xi \in \mathbb{R}^p$  y  $\eta \in \mathbb{R}^q$  dos vectores aleatorios dependientes. Se exige establecer la dependencia entre  $\xi$  y  $\eta$ , a partir de los resultados de observación  $(x_1, x_2, \dots, x_n)$  para el vector  $\xi$  y de la observación  $(y_1, y_2, \dots, y_n)$  del vector  $\eta$ .

Sea  $m(y_k, y_l)$  alguna métrica en  $\mathbb{R}^q$ . Se debe hallar una función

$$f(\cdot) : \mathbb{R}^p \rightarrow \mathbb{R}^q$$

para la cual  $E\{m(\eta, f(\xi))\}$  toma su mínimo valor. En principio, se puede escoger la norma Euclídea, descrita en (1.2), tal que

$$m(y_k, y_l) = \|y_k - y_l\|_2 \text{ en } \mathbb{R}^q$$

por lo que la solución se da en términos de la *función teórica de regresión*  $y = f(x)$ , descrita como

$$f(x) = E\{\eta | \xi = x\} \quad (2.76)$$

donde  $x$  se denomina la *variable de regresión*, mientras  $y$  es la *respuesta*.

La solución de (2.76) exige el conocimiento, por lo menos a nivel de estimación, sobre la función de densidad de probabilidad conjunta de los vectores  $\xi$  y  $\eta$ , lo cual en la práctica no es real. En este sentido, se pueden tomar las siguientes aproximaciones de solución:

- En calidad de estimación de la función teórica de regresión,  $y = f(x)$ , se define una función de aproximación de la clase  $\tilde{f} = \{f : \mathbb{R}^p \rightarrow \mathbb{R}^q\}$
- Se asume que la función  $f(x) = f(x, \theta)$ , de manera unívoca se determina por una cierta cantidad de parámetros  $\theta \in \mathbb{R}^m$ , siendo  $m$  un valor fijo. Por cierto, se asume además la dependencia lineal entre  $f(x, \theta)$  y  $\theta$ .

En consideración con lo anterior, resulta el siguiente modelo lineal de regresión:

$$y = \theta_0 f_0(x) + \theta_1 f_1(x) + \dots + \theta_{m-1} f_{m-1}(x) \quad (2.77)$$

donde  $f_i(x)$  son funciones conocidas que se determinan por la naturaleza del experimento,  $m$  es el *orden del modelo de regresión* y  $\theta = (\theta_0, \theta_1, \dots, \theta_{m-1})$  corresponde al vector de parámetros a ser estimado sobre la base de los resultados de las observaciones  $(x_1, x_2, \dots, x_n)$  y  $(y_1, y_2, \dots, y_n)$ , teniendo como restricción la desigualdad  $m < n$ .

Los siguientes modelos son los que mayor aceptación tienen en el análisis de regresión estadística:

1. *Modelos de aproximación empíricos* [17, 19].

$$\begin{aligned}
 y &= \theta_0 + \theta_1 x, & y &= \sum_k \theta_k x^k, & y &= \theta_0 + \theta_1 \ln x \\
 y &= \theta_0 + \theta_1 x + \theta_2 \frac{1}{x}, & \ln y &= \sum_k \theta_k x^k, & \ln y &= \theta_0 + \theta_1 \ln x, \\
 1/y &= \sum_k \theta_k x^k, & y &= \theta_0 + \theta_1 / \sqrt{x} & y &= \theta_0 + \theta_1 10^{\theta_2} \ln x
 \end{aligned}$$

2. *Funciones de descomposición lineal*. Usualmente, se emplean las funciones ortogonales en la representación del tipo (1.4), particularmente, las funciones de Chebyshev.

3. *Funciones de ajuste (spline)*. Cuando el modelo de regresión se da por intervalos mediante polinomios de interpolación o alguna función con estructura a priori dada.

El modelo lineal es, tal vez, el de más amplio empleo en los métodos de regresión estadística:

$$\mathbf{y} = \boldsymbol{\theta} \mathbf{x} + \boldsymbol{\varepsilon} \quad (2.78)$$

siendo  $\boldsymbol{\varepsilon}$  el vector aleatorio de errores de observación. En general, se supone que no hay errores sistemáticos de medida y, por lo tanto, el valor medio de la variable aleatoria es  $E\{\boldsymbol{\varepsilon}_k\} = 0$ , para  $k = 1, \dots, n$ , con potencia desconocida de error de observación, definida como  $E\{\boldsymbol{\varepsilon}_k, \boldsymbol{\varepsilon}_l\} = \sigma_\varepsilon^2 \delta_{kl}$ .

En caso de asumir que los errores de medición les corresponde la correlación expresada en la forma  $E\{\boldsymbol{\varepsilon}, \boldsymbol{\varepsilon}^\top\} = \sigma_\varepsilon^2 \boldsymbol{\Sigma}$ , siendo conocida la matriz  $\boldsymbol{\Sigma}$ , se puede realizar el cambio de variables:  $\boldsymbol{\lambda}' = \boldsymbol{\lambda}^{-1/2} \boldsymbol{\lambda}$ , donde  $\boldsymbol{\lambda} = \{\mathbf{y}, \mathbf{x}, \boldsymbol{\varepsilon}\}$ , que conlleva al modelo con momentos en la forma  $E\{\boldsymbol{\varepsilon}'\} = 0$  y  $E\{\boldsymbol{\varepsilon}', \boldsymbol{\varepsilon}'^\top\} = \sigma_\varepsilon^2 \mathbf{I}$ , donde  $\mathbf{I}$  es la matriz unidad, en este caso con dimensión  $n \times n$ .

Si en el modelo (2.78)  $q = 1$  y el orden de regresión  $m$  es conocido, la estimación  $\tilde{\boldsymbol{\theta}}$  del vector de parámetros  $\boldsymbol{\theta}$  del modelo puede realizarse por diferentes métodos. Sin embargo, el más empleado corresponde al *método de mínimos cuadrados*:

$$\min_{\boldsymbol{\theta}} \sum_{k=1}^n \left[ y_i - \sum_{l=0}^{m-1} f_l(x_k) \theta_l \right]^2 = \sum_{k=1}^n \left[ y_i - \sum_{l=0}^{m-1} f_l(x_k) \tilde{\theta}_l \right]^2 \quad (2.79)$$

Sea el hiperespacio lineal en  $\mathbb{R}^m$  conformado por los valores del parámetro  $\boldsymbol{\theta}$ , cuya estimación  $\tilde{\boldsymbol{\theta}}$  es obtenida por (2.79), entonces se tendrá la siguiente interpretación geométrica:

- si  $\boldsymbol{\theta} \in \mathbb{R}^m$  y  $\text{rank}\{\mathbf{x}\} = m$ , entonces el vector  $\mathbf{x}\tilde{\boldsymbol{\theta}}$  corresponde a la proyección del vector de observaciones  $\mathbf{y} \in \mathbb{R}^n$ ,  $m < n$  en el hiperespacio  $\boldsymbol{\theta} \in \mathbb{R}^m : \mathbf{x}\tilde{\boldsymbol{\theta}} = \text{pr}_{\boldsymbol{\theta}} \mathbf{y}$ . La estimación  $\tilde{\boldsymbol{\theta}}$  será única.
- si  $\boldsymbol{\theta} \in \mathbb{R}^m$ ,  $\text{rank}\{\mathbf{x}\} < m$  y el vector  $\text{pr}_{\boldsymbol{\theta}} \mathbf{y}$  no pertenece al espacio de valores de la transformación lineal  $\mathcal{G}\{\mathbf{x}\}$ , entonces el vector  $\mathbf{x}\tilde{\boldsymbol{\theta}}$  corresponde a la proyección del

vector  $\text{pr}_\theta \mathbf{y}$ . La estimación  $\tilde{\theta}$  será única.

- si  $\theta \in \mathbb{R}^m$ ,  $\text{rank}\{\mathbf{x}\} < m$  y  $\text{pr}_\theta \mathbf{y} \in \mathcal{G}\{\mathbf{x}\}$ , entonces el mínimo en (2.79) se obtiene en todo vector  $\tilde{\theta} = \tilde{\theta}' + \mathbf{h}$ , donde  $\tilde{\theta}'$  es cualquier vector del hiperespacio  $\theta \in \mathbb{R}^m$ , ortogonal a  $\mathcal{G}\{\mathbf{x}\}$ . La estimación  $\tilde{\theta}$  por mínimos cuadrados no es única.

El empleo del método de mínimos cuadrados para la obtención de las estimaciones puntuales de los parámetros desconocidos de los modelos lineales de regresión, a diferencia de otros métodos de estimación, dados en el numeral §2.2.1, no exige información a priori sobre el tipo de FDP, la cual muchas veces no se tiene en las etapas iniciales de proceso de datos.

---

**Ejemplo 2.23.** Sea la relación  $\mathbf{y} = f(\mathbf{x})$ , que deber ser confirmada a partir de un conjunto dado de  $n$  observaciones, tales que cada elemento  $y_i$  de la función  $\mathbf{y}$  o *respuesta* es obtenido para un valor fijo  $x_i$ , correspondiente a la variable  $\mathbf{x}$  o *factor*. En el caso de aceptar la hipótesis de dependencia lineal o *análisis de regresión lineal*, se deben hallar los parámetros de la recta  $\mathbf{y} = a + b\mathbf{x}$ , que para un criterio de error de representación dado, generalmente el error cuadrático medio (2.79), se alcance el valor mínimo (principio de Lagrange):

$$\min \sum_{i=1}^n (y_i (a - bx_i))^2, \mathbf{x} \in (\mathbf{x}_{\min}, \mathbf{x}_{\max}), \mathbf{y} \in (\mathbf{y}_{\min}, \mathbf{y}_{\max})$$

De tal manera, que la recta obtenida permite, con cierto valor de probabilidad, pronosticar el valor de  $\mathbf{y}$  para un valor de  $\mathbf{x}$ , dadas las observaciones  $\{x_i, y_i\}$ .

Sean los resultados de medida  $y_i$  distribuidos normalmente,  $p(\mathbf{y}) \sim \mathcal{N}(\sigma_{\mathbf{y}}, y_i)$ , tal que la probabilidad de obtener el valor  $y_i$  sea igual a:

$$P_{a,b}(y_i) \sim \frac{1}{\sigma_{\mathbf{y}}} \exp(-y_i - a - bx_i)^2 / 2\sigma_{\mathbf{y}}^2$$

la cual depende de los coeficientes  $a$  y  $b$ . Por lo tanto, la probabilidad de obtener el conjunto de resultados de medida  $y_1, \dots, y_n$ , asumiendo su independencia estadística, es igual a

$$P_{a,b}(y_1, \dots, y_n) = P_{a,b}(y_1) \dots P_{a,b}(y_n) \sim \frac{1}{\sigma_{\mathbf{y}}^n} \exp(-\chi^2/2)$$

siendo  $\chi^2 = \sum_{i=1}^n (y_i - a - bx_i)^2 / \sigma_{\mathbf{y}}^2$ .

La estimación de las constantes  $a$  y  $b$ , por el método de máximo de probabilidad a posteriori corresponde a la mayor probabilidad de  $P_{a,b}(y_1, \dots, y_n)$ , esto es, cuando la suma  $\chi^2$  es mínima o método de *mínimos cuadrados*, descrito por la expresión (2.79). El valor óptimo se halla diferenciando parcialmente la suma por las variables  $a$  y  $b$ , e igualando a cero:

$$\begin{aligned} \frac{\partial \chi^2}{\partial a} &= \left(-2/\sigma_{\mathbf{y}}^2\right) \sum_{i=1}^n (y_i - a - bx_i) = 0 \\ \frac{\partial \chi^2}{\partial b} &= (-2\sigma_{\mathbf{y}}^2) \sum_{i=1}^n x_i (y_i - a - bx_i) = 0 \end{aligned}$$



Las anteriores expresiones se pueden representar como un sistema de ecuaciones, teniendo como variables  $a$  y  $b$  (ecuaciones normales), de la forma:

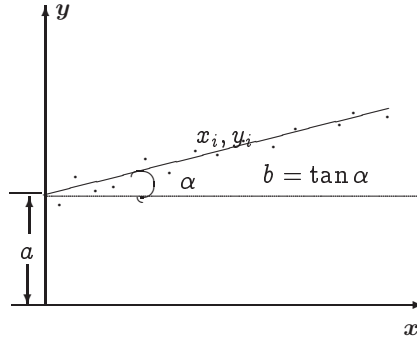
$$\begin{aligned} an + b \sum_{i=1}^n x_i &= \sum_{i=1}^n y_i \\ a \sum_{i=1}^n x_i + b \sum_{i=1}^n x_i^2 &= \sum_{i=1}^n x_i y_i \end{aligned}$$

por lo que se obtiene ( $\Sigma$  denota la suma por  $i = 1, \dots, n$ ):

$$a = \frac{(\Sigma x_i^2)(\Sigma y_i) - (\Sigma x_i)(\Sigma x_i y_i)}{\Delta} \quad (1a)$$

$$b = \frac{n(\Sigma x_i y_i) - (\Sigma x_i)(\Sigma y_i)}{\Delta} \quad (1b)$$

siendo  $\Delta = n(\Sigma x_i^2) - (\Sigma x_i)^2$ .



**Figura 2.10.** Interpretación geométrica de la regresión lineal

Geoméricamente, el coeficiente  $a$ , obtenido en (1a), corresponde a la distancia desde el origen de las coordenadas hasta el punto, donde se intercepta la línea de regresión con la ordenada, mientras el coeficiente  $b$ , obtenido en (1b), corresponde a la tangente del ángulo  $\alpha$  de inclinación de la línea de regresión, como se muestra en la Figura 2.10.

**Ejemplo 2.24.** Realizar la prueba de hipótesis sobre la validez de ecuación de regresión lineal simple obtenida en el ejemplo 2.23.

Se asume que el valor que pondera qué tanto el conjunto de resultados de medida, contenidos en las parejas  $\{(x_i, y_i) : i = 1, \dots, n\}$ , confirma la hipótesis sobre la dependencia lineal entre  $\mathbf{x}$  e  $\mathbf{y}$ , corresponde al coeficiente de correlación dado en (2.46),

$$\rho = \frac{\sigma_{xy}}{\sigma_x \sigma_y} \quad (2)$$

donde la suma,  $\sigma_{xy} = \frac{1}{n} \sum_{i=1}^n (x_i - \tilde{m}_{1x})(y_i - \tilde{m}_{1y})$  es un término paralelo al segundo momento centralizado de una variable dado en (2.9), definida como la *desviación combinada*, pero para la cual no se cumplen todas las condiciones (2.14). Reemplazando la anterior definición en (2) y teniendo en cuenta la estimación de la varianza dada en (2.73), se obtiene la

estimación del coeficiente de correlación lineal en la forma

$$\tilde{\rho} = \frac{\sum_i (x_i - \tilde{m}_{1x})(y_i - \tilde{m}_{1y})}{\left( \sum_i (x_i - \tilde{m}_{1x})^2 \sum_i (y_i - \tilde{m}_{1y})^2 \right)^{1/2}} \quad (3)$$

Por cuanto,  $-1 \leq \rho \leq 1$ , si  $|\tilde{\rho}| \rightarrow 1$ , entonces los puntos están ubicados en las cercanías de una línea recta. En cambio, si  $|\tilde{\rho}| \rightarrow 0$ , los puntos no están correlacionados y, por lo tanto, no se agrupan sobre ninguna recta. A partir de coeficiente de correlación (3), el modelo de la regresión ortogonal se construye por la expresión:

$$\mathbf{y} = \tilde{m}_{1y} + \frac{2\tilde{\rho}}{\sigma_0 + \sqrt{\sigma_0^2 + 4\tilde{\rho}^2}} (\mathbf{x} - \tilde{m}_{1x})$$

la cual está directamente relacionada con el concepto de la elipse de dispersión, mostrada en la Figura 2.8(b).

La estimación de los coeficientes de regresión es analizada estadísticamente, mediante la prueba de hipótesis  $H_0 : \rho = 0$ , esto es, si hay diferencia o no, desde el punto de vista estadístico, en la estimación del coeficiente de regresión con respecto al valor 0 [19]. Se puede demostrar que cuando  $\rho = 0$ , la estadística  $\tilde{\rho}$  se describe mediante la  $\Gamma$ -densidad (2.28), de la forma [20]:

$$\frac{\Gamma\{\frac{1}{2}(n-1)\}}{\Gamma\{\frac{1}{2}(n-2)\}\sqrt{\pi}} (1 - \tilde{\rho}^2)^{(n-4)/2}$$

que es equivalente a la distribución de Student (2.30) con  $n - 2$  grados de libertad para la variable aleatoria

$$(n-2)^{1/2} \frac{\tilde{\rho}}{\sqrt{(1 - \tilde{\rho}^2)}}$$

El caso más importante de análisis es la prueba de hipótesis  $H_0 : \rho = 0$  contra  $H_1 : \rho \neq 0$ . La hipótesis  $H_0$  se rechaza con un nivel de significación  $\alpha$ . Por lo anterior, los límites del intervalo de análisis se fijan:

$$\hat{t} = \left| \tilde{b} \right| / \tilde{\sigma}_b \geq t_{1-\frac{1}{2}\alpha} \{n-2\} \quad (4)$$

siendo  $\tilde{b}$  la estimación del coeficiente de regresión. Si la desigualdad (4) se cumple, entonces se puede decir que  $\tilde{\rho}$  significativamente se diferencia de cero.

La estimación del coeficiente de correlación (prueba de existencia sobre la correlación), se realiza de manera similar:

$$\hat{t} = \frac{\tilde{\rho}\sqrt{n-2}}{1 - \tilde{\rho}^2} \geq t_{1-\frac{1}{2}\alpha} \{n-2\}$$

Si la anterior condición se cumple, entonces la hipótesis  $H_0 : \tilde{\rho} = 0$ , se rechaza.

La prueba completa sobre la validez de ecuación de regresión, se efectúa mediante la  $F$ -estadística

de Fisher (2.32):

$$\tilde{F} = \frac{\frac{1}{k-2} \sum_{i=1}^k n_i (\tilde{m}_{1y_i} - \tilde{y}_i)^2}{\frac{1}{n-k} \sum_{i=1}^k \sum_{l=1}^{n_i} (y_{il} - \tilde{m}_{1y})^2} \quad (5)$$

donde  $\tilde{y}_i = a + b\tilde{x}_i$ . El criterio implica que si la suma de los valores medios de las desviaciones de grupo de la regresión directa, dividida sobre la suma de las desviaciones de los valores  $\mathbf{y}$  con respecto a los valores medios de grupo, alcanza o excede los límites de significación, entonces la hipótesis sobre linealidad hay que rechazarla. La estimación (5) exige que la longitud del vector  $\mathbf{y}$  sea mayor, en  $k$  valores a la respectiva longitud del vector  $\mathbf{x}$ , esto es, que a cada valor  $x_i$  le corresponde  $n_i$  valores del vector  $\mathbf{y}$ , lo que en la práctica, frecuentemente ocurre.

Si la hipótesis sobre la linealidad no puede ser tomada, entonces se pueden realizar pruebas sobre modelos no lineales, en primera instancia, de tipo polinomial:

$$y = a + bx + cx^2 + \dots$$

Aunque su implementación exige el empleo de métodos computacionales. Sin embargo, después de cierto orden, la varianza empieza a crecer desmesuradamente, por lo que el algoritmo no siempre converge.

En general, empleando el método de los mínimos cuadrados (2.79), se puede construir, prácticamente, cualquier forma de dependencia no lineal. En este caso, se emplea alguna transformación de linealización, por cuanto, solamente las funciones lineales pueden ser reconstruidas por el método de mínimos cuadrados.

---

## Problemas

**Problema 2.12.** Dada una observación de valores independientes con volumen  $n$ , constatar que la estimación del parámetro  $\lambda$  de la FDP exponencial, calculada por cualquiera de los métodos de los momentos o de máximo verosimilitud, converge en el mismo valor. Calcular la varianza de la estimación y analizar su efectividad.

**Problema 2.13.** Dado el valor del coeficiente de confianza,  $1 - \alpha = .95$ , hallar los valores extremos del intervalo de confianza del parámetro  $\lambda$  de una variable aleatoria con FDP exponencial, dada una observación independiente de valores con volumen  $n$ .

**Problema 2.14.** Dada una observación de valores independientes con volumen  $n$ , mediante el método de máxima verosimilitud hallar la estimación del valor medio de una variable aleatoria con FDP Rayleigh, descrita por (1) en el ejemplo 2.15.

**Problema 2.15.** Probar la hipótesis nula  $H_0$  sobre el valor de la varianza,  $\sigma_\xi^2 = k_0$ , de una variable aleatoria Gaussiana, con volumen  $n$  y valor  $m_{1\xi}$ , contra la hipótesis alternativa  $H_1$  sobre el valor  $\sigma_\xi^2 = k_1$ , siendo  $k_0 < k_1$ .

**Problema 2.16.** Basados en el criterio de Neyman-Pearson realizar la prueba de hipótesis sobre el tipo de FDP con parámetros conocidos, para un valor dado de significación  $\alpha$ . Sea la hipótesis nula,  $p(x|H_0) = (2\pi)^{-1/2} \exp(-x^2/2)$ , la FDP normal, mientras la hipótesis alternativa corresponde a la densidad exponencial simétrica  $p(x|H_1) = (1/2) \exp(-|x|)$ .

**Problema 2.17.** Un dosímetro, durante un tiempo  $T = 10s$ , registra la cantidad  $k = 20$  partículas. Realizar la prueba de hipótesis mediante el criterio de Neyman-Pearson para un valor dado de significación  $\alpha$ , siendo la hipótesis nula  $H_0$ , que considera que el resultado de la observación está condicionado por la radiación de fondo, contra la hipótesis alternativa  $H_1$  que corresponde a la suma del fondo mas una fuente de radiación. Los valores medios de partículas por segundo son conocidos e iguales a  $n_0 = 0.2$  y  $n_1 = 0.8$ , de forma correspondiente.

**Problema 2.18.** Probar la hipótesis nula  $H_0$  sobre el valor del parámetro  $\lambda = k_0$  de una variable aleatoria con FDP exponencial con volumen  $n$ , contra la hipótesis alternativa  $H_1$  sobre el valor  $\lambda = k_1$ , siendo  $k_0 < k_1$ .

**Problema 2.19.** Dadas dos observaciones independientes de variables aleatorias Gaussianas, con volúmenes  $n_1 = 16$  y  $n_2 = 25$ , se obtiene las correspondientes estimaciones del coeficiente de correlación  $r_1 = 0.5$  y  $r_2 = 0.7$ . Realizar la prueba de hipótesis, con valor de significación  $\alpha = 0.05$ , por la cual ambas observaciones corresponden a un mismo conjunto general.

**Problema 2.20.** Estimar los coeficientes de la regresión lineal  $y = \tilde{a} + \tilde{b}x$ , dadas las observaciones,

$x_i$	6.4	8.0	7.4	6.8	7.2	6.7	7.7	8.2
$y_i$	2.2	5.2	4.8	3.3	4.0	3.0	4.2	5.3

Realizar la prueba de hipótesis sobre la validez de ecuación de regresión lineal simple obtenida y estimar el intervalo de confianza para un valor dado  $1 - \alpha = 0.9$ .

## 2.3. Teoría de decisión estadística

La estimación óptima de procesos aleatorios está basada en la teoría de decisión estadística, en la medida en que la información extraída del proceso aleatorio se emplea para la obtención de algún objetivo de análisis, cuyo resultado es una decisión.

Aunque la variedad de tareas, en las cuales se hace necesario la toma de decisión, es amplia se tiene una formulación matemática con las siguientes propiedades:

- (a). Cualquier solución está orientada a la obtención de un objetivo concreto, que bien puede estar definido por una cantidad, finita o infinita, de variantes con sus respectivas bondades y restricciones, sobre las cuales se estima la calidad de la decisión a tomar.
- (b). La decisión se toma, basados tanto en la información como en la evidencia estadística a priori y obtenidos durante el proceso de solución.
- (c). La toma de decisión dentro de un conjunto de posibles variantes puede contemplar la incertidumbre, determinada por la respectiva probabilidad de toma de cada decisión.

La decisión seleccionada dentro de todas las posibles variantes, que brinde la mayor bondad de solución de acuerdo a un objetivo propuesto, se denomina *óptima*.

### 2.3.1. Definiciones básicas

**Funciones de decisión.** Sea la tarea de estimación de procesos aleatorios, en la cual se tiene la trayectoria de un vector aleatorio  $\mathbf{y}(t) \in \mathfrak{Y}$ . Así mismo, sea  $\mathbf{x}(t) \in \mathfrak{X}$  un proceso aleatorio vectorial, cuyas trayectorias no son conocidas al observador y que corresponden al valor verdadero del proceso, pero cuya información está contenida en los resultados de observación  $\mathbf{y}(t)$ .

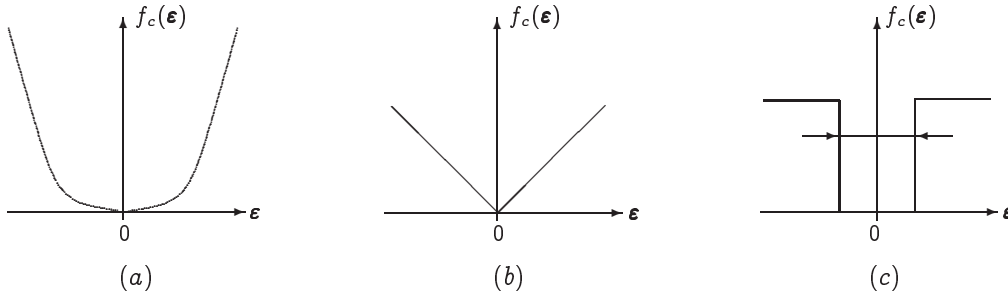
Sea el vector  $\mathbf{D} \in \mathfrak{D}$  de cada posible solución, que se puede tomar con respecto a  $\mathbf{x}(t)$  por los resultados de observación  $\mathbf{y}(t)$ , esto es, cada solución establece una relación entre los puntos de  $\mathfrak{Y}$ . Sea  $\mathbf{g} \in \mathcal{G}$  la transformación de la forma  $\mathbf{D} = \mathbf{g}(\mathbf{y})$ .

Las funciones de decisión pueden ser, tanto determinísticas cuando la transformación no contiene ninguna medida de incertidumbre, como randomizada cuando la transformación implica que para cada  $\mathbf{y}$ , cualquier  $\mathbf{D}$  se toma con una medida de probabilidad condicional  $p(\mathbf{D}|\mathbf{y})$ .

**Funciones de pérdida.** Como resultado de la toma de una u otra decisión pueden ocurrir errores, por lo cual se establece una medida cuantitativa de la respectiva ganancia o pérdida, en forma de dependencia  $f_c\{\mathbf{x}, \mathbf{d}\}$  denominada *función de pérdida*, que se selecciona en concordancia con la naturaleza física de la tarea a resolver. Debido a que la solución  $\mathbf{D} = \mathbf{g}(\mathbf{y})$  depende de la trayectoria del proceso aleatorio  $\mathbf{y}(t)$ , entonces el valor actual de pérdida  $f_c\{\mathbf{x}, \mathbf{g}(\mathbf{y})\}$  es también aleatorio. Por lo tanto, las funciones de decisión es debido seleccionarlas de acuerdo a la comparación de las características de aleatoriedad de la función de pérdida. Así por ejemplo, para el caso de la tarea de estimación, la decisión  $\mathbf{D}$

representa la estimación  $\tilde{\mathbf{x}}(t)$  que corresponde al vector proceso aleatorio  $\mathbf{x}(t)$ . Al introducir la medida de error  $\boldsymbol{\varepsilon}(t) = \mathbf{x}(t) - \tilde{\mathbf{x}}(t)$ , entonces, la función de pérdida se describe como  $f_c\{\mathbf{x}, \tilde{\mathbf{x}}\} \triangleq f_c\{\boldsymbol{\varepsilon}\}$ . Por cierto, a la función de pérdidas se denomina *permitida* si presenta las siguientes propiedades:

- (a).  $f_c\{\boldsymbol{\varepsilon}\}$  es función escalar de  $n$  variables, correspondientes a la dimensión del proceso  $\mathbf{x}(t)$ ,
- (b).  $f_c\{\boldsymbol{\varepsilon} = 0\} = 0$ ,
- (c).  $f_c\{\boldsymbol{\varepsilon}\} = f_c\{-\boldsymbol{\varepsilon}\}$ , la función de pérdida es simétrica, en el sentido en que no importa en que dirección se tenga el error, la pérdida se considera la misma.
- (d).  $f_c\{\boldsymbol{\varepsilon}_m\} > f_c\{\boldsymbol{\varepsilon}_n\}$ ,  $\forall \rho(\boldsymbol{\varepsilon}_m) \geq \rho(\boldsymbol{\varepsilon}_n)$ , donde  $\boldsymbol{\varepsilon}_m$  y  $\boldsymbol{\varepsilon}_n$  son los valores de error de estimación,  $\rho$  es una función escalar positiva semidefinida convexa, con argumento de  $n$  variables, y corresponde a la medida de distancia del error  $\boldsymbol{\varepsilon}$ , desde el origen de las coordenadas de un espacio euclídeo con dimensión  $n$ . La función  $f_c\{\boldsymbol{\varepsilon}\}$  es monótona ascendente con respecto al incremento del argumento  $\boldsymbol{\varepsilon}$ .



**Figura 2.11.** Funciones de pérdida.

En la práctica, amplio uso tienen las siguientes tres formas de funciones permitidas de pérdida:

a. Cuadrática, Figura 2.11(a),  $f_c(\boldsymbol{\varepsilon}) = \|\boldsymbol{\varepsilon}\|_{\mathbf{B}}^2$  (2.80a)

b. Por módulo, Figura 2.11(b),  $f_c(\boldsymbol{\varepsilon}) = \|\boldsymbol{\varepsilon}\|_{\mathbf{B}}$  (2.80b)

c. Simple, Figura 2.11(c),  $f_c(\boldsymbol{\varepsilon}) = \begin{cases} 0, & \|\boldsymbol{\varepsilon}\|_{\mathbf{B}} < e/2 \\ 1/e, & \|\boldsymbol{\varepsilon}\|_{\mathbf{B}} \geq e/2 \end{cases}$  (2.80c)

donde  $\|\cdot\|$  es la norma del vector,  $e$  es una constante y  $\mathbf{B}$  es una matriz positiva semidefinida de dimensión  $n \times n$ , compuesta por los elementos  $b_{ij}$  que son factores de peso. En la estimación de señales es usual el empleo de la norma del error

$$\|\boldsymbol{\varepsilon}_{\mathbf{B}}\|^2 = \boldsymbol{\varepsilon}^T \mathbf{B} \boldsymbol{\varepsilon}$$

**Funciones de riesgo.** Como anteriormente se dijo, los valores concretos de la función de pérdida  $f_c(\mathbf{x}, \mathbf{D})$  son aleatorios, y por lo tanto, las funciones de decisión se seleccionan basados en la comparación de las propiedades estadísticas de la función de pérdida, la cual se denomina *función de riesgo* (o riesgo, simplemente) y corresponde a las pérdidas que en promedio se obtienen sobre el conjunto de decisiones posibles.

La selección de la decisión óptima se lleva a cabo mediante la minimización de los riesgos por todas las posibles funciones de decisión. En la teoría de toma de decisiones, en calidad de criterio de calidad generalizado, se emplea el riesgo medio, que para el caso de las funciones de decisión *randomizadas* corresponde a la esperanza matemática de la función de pérdidas en la forma,

$$R = E \{f_c\{\mathbf{x}, \mathbf{D}\}\} = \iiint f_c\{\mathbf{x}, \mathbf{D}\} p(\mathbf{D}|\mathbf{y}) p(\mathbf{x}, \mathbf{y}) d\mathbf{D} d\mathbf{y} d\mathbf{x} \quad (2.81)$$

donde  $p(\mathbf{D}|\mathbf{y})$  es la densidad de probabilidad condicional de la toma de decisión  $\mathbf{D}$  para un valor dado del vector  $\mathbf{y}$ , que caracteriza la función de decisión,  $p(\mathbf{x}, \mathbf{y})$  es la FDP conjunta de los vectores  $\mathbf{x}$  e  $\mathbf{y}$ . Teniendo en cuenta que

$$p(\mathbf{x}, \mathbf{y}) = p(\mathbf{x})p(\mathbf{y}|\mathbf{x}) = p(\mathbf{y})p(\mathbf{x}|\mathbf{y})$$

entonces, el riesgo medio (2.81) se puede representar en forma alterna como

$$R = \iiint f_c(\mathbf{x}, \mathbf{D}) p(\mathbf{D}|\mathbf{y}) p(\mathbf{x}, \mathbf{y}) p(\mathbf{y}) d\mathbf{D} d\mathbf{y} d\mathbf{x} \quad (2.82a)$$

$$R = \iiint f_c(\mathbf{x}, \mathbf{D}) p(\mathbf{D}|\mathbf{y}) p(\mathbf{y}, \mathbf{x}) p(\mathbf{x}) d\mathbf{D} d\mathbf{y} d\mathbf{x} \quad (2.82b)$$

donde  $p(\mathbf{x}|\mathbf{y})$  es la FDP condicional del proceso no observado  $\mathbf{x}(t)$  para un valor dado  $\mathbf{y}(t)$ , además,  $p(\mathbf{x})$  y  $p(\mathbf{y})$ , son la FDP marginales de  $\mathbf{x}$  e  $\mathbf{y}$ , respectivamente, mientras,  $p(\mathbf{y}|\mathbf{x})$  es la FDP condicional del proceso observado  $\mathbf{y}$  para una valor de  $\mathbf{x}$ , que es la función de verosimilitud de  $\mathbf{x}$ .

Las expresiones (2.81), (2.82a), y (2.82b) caracterizan las pérdidas en promedio por todas las posibles decisiones  $\mathbf{D} \in \mathfrak{D}$ , así como todos los posibles valores de  $\mathbf{x}$  e  $\mathbf{y}$ . La optimización de la función de decisión consiste en la selección de aquella función  $p(\mathbf{D}|\mathbf{y})$  que brinde el mínimo riesgo medio, determinado por la función (2.81).

Cuando se analizan funciones *no randomizadas* de decisión, a cada  $\mathbf{y}$  le corresponde una determinada solución  $\mathbf{D} = \mathbf{g}(\mathbf{y})$  y la FDP condicional es  $p(\mathbf{D}|\mathbf{y}) = \delta(\mathbf{D} - \mathbf{g}(\mathbf{y}))$ , con lo cual el riesgo medio es igual a,

$$R_g = E \{f_c\{\mathbf{x}, \mathbf{g}(\mathbf{y})\}\} = \iint f_c\{\mathbf{x}, \mathbf{g}(\mathbf{y})\} p(\mathbf{x}, \mathbf{y}) d\mathbf{y} d\mathbf{x} \quad (2.83)$$

En general, el riesgo medio (2.83) es una función no lineal con relación a la función de decisión  $\mathbf{g}(\mathbf{y})$ . Las respectivas funciones (2.82a) y (2.82b) del riesgo medio para el caso de

las funciones de decisión no randomizadas tienen la forma,

$$R_g = \iint f_c(\mathbf{x}, \mathbf{g}(\mathbf{y}))p(\mathbf{x}|\mathbf{y})p(\mathbf{y})d\mathbf{y}d\mathbf{x} \quad (2.84a)$$

$$R_g = \iint f_c(\mathbf{x}, \mathbf{g}(\mathbf{y}))p(\mathbf{y}|\mathbf{x})p(\mathbf{x})d\mathbf{y}d\mathbf{x} \quad (2.84b)$$

de los cuales se puede determinar el riesgo condicional  $r_g(\mathbf{x})$  por  $\mathbf{x}$ , que corresponde al promedio de la función de pérdida por la función de verosimilitud  $p(\mathbf{y}|\mathbf{x})$  para un valor dado de  $\mathbf{x}$ ,

$$r_g(\mathbf{x}) = \int f_c(\mathbf{x}, \mathbf{g}(\mathbf{y}))p(\mathbf{y}|\mathbf{x})d\mathbf{y} \quad (2.85)$$

El promedio del riesgo condicional  $r_g(\mathbf{x})$  por todos los posibles valores de  $\mathbf{x}$ , de acuerdo a (2.84b), resulta en el riesgo medio,

$$R_g = \int r_g(\mathbf{x})p(\mathbf{x})d\mathbf{x} \quad (2.86)$$

El riesgo condicional  $r_g(\mathbf{x})$  caracteriza la calidad del sistema que emplea la función de decisión  $\mathbf{g}(\mathbf{y})$ , para un valor dado de  $\mathbf{x}$ . Así que se prefiere aquella función de decisión  $\mathbf{g}(\mathbf{y})$  que presente el menor valor de  $r_g(\mathbf{x})$ . Se define la función de decisión  $\mathbf{g}_1(\mathbf{y})$  como *uniformemente mayor* que  $\mathbf{g}_2(\mathbf{y})$ , cuando se cumple que  $r_{g_1}(\mathbf{x}) \leq r_{g_2}(\mathbf{x}) \forall \mathbf{x} \in \mathfrak{X}$ , además,  $r_{g_1}(\mathbf{x}) < r_{g_2}(\mathbf{x})$  por lo menos para un valor de  $\mathbf{x}$  [21]. De otra parte, la función decisión  $\mathbf{g}$  se denomina *permitida*, si en  $\mathfrak{G}$  no existe otra función decisión  $\mathbf{g}_1$ , que sea uniformemente mayor que  $\mathbf{g}$ .

La clase  $\mathfrak{G}$  de funciones de decisión se denomina *completa*, si para cualquier  $\mathbf{g}_i \notin \mathfrak{G}$  se halla una función de decisión  $\mathbf{g} \in \mathfrak{G}$  uniformemente mayor que  $\mathbf{g}_i$ . En este sentido, si se tiene una clase completa  $\mathfrak{G}$  de funciones de decisión completa, entonces la función de decisión preferible es suficiente con esta clase sin necesidad de analizar las funciones  $\mathbf{g} \notin \mathfrak{G}$  y, por lo tanto, la construcción de clases completas es una tarea importante en la teoría de decisión estadística.

Cuando se tiene alguna función de decisión  $\mathbf{g}(\mathbf{y})$ , que minimice el riesgo condicional  $r_G(\mathbf{x})$ , es suficientemente diferente para los valores de  $\mathbf{x}$ . Por esta razón, la decisión óptima se selecciona basados en diferentes aproximaciones, entre ellas, las de Bayes, minimax, máximo de verosimilitud, etc.

Cuando las observaciones son un conjunto de valores aleatorios, o bien son procesos continuos sobre un intervalo de tiempo de análisis, pero con dimensión alta o infinita, entonces en la teoría de decisión estadística se emplea el concepto de *estadísticas suficientes*, que corresponde a la clase de transformaciones  $\gamma(\mathbf{y})$ , que contienen toda la información, considerada necesaria para la toma de decisión y que está inmersa en las observaciones  $\mathbf{y} \in \mathfrak{Y}$ . Para estas transformaciones la función de decisión óptima  $\mathbf{g}(\gamma(\mathbf{y}))$ , que depende de  $\gamma$ , brinda el mismo valor de riesgo que la decisión óptima  $\mathbf{g}(\mathbf{y})$ . En este caso, la transformación  $\gamma(\mathbf{y})$ , que es la suficiente estadística, presenta importancia para hallar las funciones de forma



de decisión óptimas del tipo, tanto bayesianas, como no bayesianas.

En el caso de la aproximación bayesiana, para la toma de decisión, las estadísticas suficientes corresponden a funciones vectoriales conocidas  $\boldsymbol{\gamma} = \{\gamma_i(\mathbf{y}) : i = 1, \dots, n\}$ , por medio de las cuales la FDP conjunta se puede representar en la forma

$$p(\mathbf{x}, \mathbf{y}) = f_c(\mathbf{y})f(\gamma_1(\mathbf{y}), \dots, \gamma_n(\mathbf{y}), \mathbf{x})$$

donde el factor de multiplicación  $f_c(\mathbf{x})$  no depende de  $\mathbf{x}$ , y la función  $f$  está relacionada con  $\mathbf{y}$ , solamente mediante las funciones  $\gamma_1(\mathbf{y}), \dots, \gamma_n(\mathbf{y})$ .

Como se observa, en este caso  $f$  tiene una cantidad finita de variables. Además, la FDP condicional a posteriori va a depender solamente de  $\gamma_i(\mathbf{y}), \dots, \gamma_n(\mathbf{y})$ ,

$$p(\mathbf{x}|\mathbf{y}) = \frac{f(\gamma_1(\mathbf{y}), \dots, \gamma_n(\mathbf{y}), \mathbf{x})}{\int f(\gamma_1(\mathbf{y}), \dots, \gamma_n(\mathbf{y}), \mathbf{x})d\mathbf{x}}$$

Por lo tanto, las funciones  $\gamma_i(\mathbf{y})$  son suficientes para el cálculo de la FDP a posteriori condicional,

$$p(\mathbf{x}|\mathbf{y}) = p(\mathbf{x}|\gamma_1(\mathbf{y}), \dots, \gamma_n(\mathbf{y}))$$

que al reemplazar en (2.84a) implica que el algoritmo óptimo de estimación también está relacionado con  $\mathbf{y}$  solamente a través de la estadística suficiente,  $\mathbf{g}(\mathbf{y}) = \mathbf{g}(\gamma_1(\mathbf{y}), \dots, \gamma_n(\mathbf{y}))$ .

En la aproximación no bayesiana para la toma de decisión mediante las estadísticas suficientes, de manera similar, la función de verosimilitud tiene la forma,  $p(\mathbf{y}|\mathbf{x}) = f_c(\mathbf{y})$ .

Una estadística suficiente se considera que es más preponderante, entre mayor valor de compresión brinde de los datos observados  $\mathbf{y}(t)$ , esto es, entre menor sea su dimensión de representación. La transformación suficiente de mínima dimensión se denomina *estadística mínima suficiente* para una clase dada de tareas. Como ejemplos de estadísticas mínimas suficientes se tiene la relación de verosimilitud en la tarea de detección binaria de señales, y el funcional de verosimilitud en la estimación de parámetros de las señales.

La tarea más importante de estimación corresponde al riesgo condicional por  $\mathbf{y}$  o riesgo a posteriori  $r_g(\mathbf{y})$ , que es la esperanza media condicional para una función de decisión dada  $\mathbf{g}$  desde la trayectoria de  $\mathbf{y}$ , mientras, el riesgo  $r_g(\mathbf{y})$  se forma por el promedio de la función de pérdida sobre toda la función de densidad de probabilidad a posteriori, para una trayectoria conocida  $\mathbf{y}$ ,

$$r_g(\mathbf{y}) = \int f_c(\mathbf{x}, \mathbf{g}(\mathbf{y}))p(\mathbf{x}|\mathbf{y})d\mathbf{x} \quad (2.87)$$

El riesgo (2.87) determina el valor medio de pérdidas como resultado de la toma de la decisión  $\mathbf{D} = \mathbf{g}(\mathbf{y}) \in \mathcal{D}$ , que corresponde a una observación para  $\mathbf{y}$ . El riesgo medio  $R_g$ , en concordancia con (2.84a) es el promedio del riesgo a posteriori  $r_g(\mathbf{y})$  por todos los posibles valores de  $\mathbf{y}$ ,

$$R_g = \int r_g(\mathbf{y})p(\mathbf{y})d\mathbf{y} \quad (2.88)$$

### 2.3.2. Decisión Bayesiana

En general, la búsqueda de la función de decisión óptima considera dos casos extremos de análisis sobre el estado de conocimiento a priori para un proceso vectorial no observado  $\mathbf{x}$ : conocimiento completo a priori, cuando se tiene la FDP,  $p(\mathbf{x})$ , y conocimiento incompleto a priori, cuando  $p(\mathbf{x})$  es totalmente desconocida. En el primer caso, se construyen las *funciones de decisión bayesiana*, en el segundo las denominadas *no bayesianas*.

En la aproximación bayesiana se cumplen las dos siguientes condiciones:

1. El proceso  $\mathbf{x}$  es aleatorio y la respectiva FDP marginal  $p(\mathbf{x})$  existe.
2. La FDP a priori  $p(\mathbf{x})$  es conocida al observador

En este caso, la decisión óptima se halla al minimizar el riesgo medio (2.83),

$$\hat{R}_g = \min_g R_g = \min_g \iint f_c(\mathbf{x}, \mathbf{g}(\mathbf{y})) p(\mathbf{x}, \mathbf{y}) d\mathbf{x} d\mathbf{y} \quad (2.89)$$

La función  $\hat{\mathbf{g}}$ , que cumple (2.89), se denomina *función de decisión bayesiana*, al igual que las correspondientes decisión y riesgo medio obtenido. De esta manera, el criterio bayesiano de optimización consiste en la minimización del riesgo medio para una FDP  $p(\mathbf{x})$ , dada a priori. Debido a que el mismo riesgo medio varía en dependencia de la estructura de función de pérdida  $f_c(\mathbf{x}, \mathbf{g}(\mathbf{y}))$ , entonces el criterio de optimización, en forma general, depende de la forma de la función de pérdida. Sin embargo, es conocido que el conjunto total de las decisiones bayesianas, con relación a todas las posibles FDP a priori  $p(\mathbf{x})$ , conforma una clase completa.

Como se observa de (2.88), el procedimiento de minimización del riesgo medio  $R_g$  por  $\mathbf{g}$  implica la minimización por  $\mathbf{g}$  del riesgo a posteriori  $r_g(\mathbf{y})$ , definido en (2.87), que está condicionado por el criterio positivo semidefinido de la FDP  $p(\mathbf{y})$ , lo que implica que el criterio de optimización bayesiano es equivalente al criterio del mínimo del riesgo a posteriori,

$$\hat{r}_g(\mathbf{y}) = \min_g r_g(\mathbf{y}) = \min_g \int f_c(\mathbf{x}, \mathbf{g}(\mathbf{y})) p(\mathbf{x}|\mathbf{y}) d\mathbf{x} \quad (2.90)$$

desde el punto de vista analítico, la tarea (2.90) en algunos casos tiene solución.

En la estimación de procesos aleatorios  $\mathbf{x}$ , la función de pérdida usualmente se determina por un error  $\boldsymbol{\epsilon}$ , cuyo valor pueda ser minimizado seleccionando la estimación  $\tilde{\mathbf{x}}$ . La estimación que cumple esta exigencia se denomina *óptima bayesiana*, o simplemente *óptima* cuando de forma tácita se asume que es bayesiana, la cual en forma general, depende de la selección de la función de pérdida  $f_c(\boldsymbol{\epsilon})$ . Sin embargo, para una clase amplia de tareas, las estimaciones, que son óptimas para una clase de función de pérdida, conservan esta propiedad para la clase completa de funciones de pérdida. Esto es, se puede hablar de invariabilidad de la estimación óptima. En particular, según teorema de Sherman [12] se tiene que si  $f_c(\boldsymbol{\epsilon})$  es una función permitida de pérdida y la FDP a posteriori del proceso  $\mathbf{x}$  es del tipo unimodal y simétrica con relación a la moda, entonces la estimación óptima

bayesiana corresponde a la esperanza matemática condicional del vector  $\mathbf{x}$ . Este caso tiene lugar, por ejemplo, en la estimación lineal de procesos Gaussianos.

---

**Ejemplo 2.25.** Sea la función cuadrática de pérdida (2.80a) y el riesgo bayesiano (2.89). Hallar el valor de la estimación  $\tilde{\mathbf{x}}$  que brinde el menor error para una observación dada  $\mathbf{y}$ .

El riesgo medio bayesiano para el criterio de error dado, tiene la forma

$$R_g = \iint (\mathbf{x} - \tilde{\mathbf{x}})^\top \mathbf{B}(\mathbf{x} - \tilde{\mathbf{x}}) p(\mathbf{x}, \mathbf{y}) d\mathbf{x} d\mathbf{y}$$

Por lo tanto, el riesgo a posteriori tiene la forma

$$r_g(\mathbf{y}) = \int (\mathbf{x} - \tilde{\mathbf{x}})^\top \mathbf{B}(\mathbf{x} - \tilde{\mathbf{x}}) p(\mathbf{x}|\mathbf{y}) d\mathbf{x} \quad (1)$$

donde  $\mathbf{B}$  es una matriz positiva definida, cuyos elementos tienen las unidades de la función de pena por error sobre las unidades del producto de los errores de las respectivas componentes del proceso vectorial estimado  $\mathbf{x}$ . Como se observa en (1), el riesgo a posteriori está relacionado con el momento central de segundo orden de la FDP a posteriori, lo que significa que el mínimo valor de posible riesgo a posteriori, en el caso de la función cuadrática de pérdida, se suple con la estimación  $\tilde{\mathbf{x}}$  con el mínimo valor de la traza de la matriz de momentos centrales escalares de segundo orden (si el proceso es escalar, entonces con el mínimo valor cuadrático medio). En este caso, el valor de la estimación que brinde el menor error cuadrático medio, dada la observación  $\mathbf{y}$ , se obtiene haciendo igual a cero el gradiente  $r_g(\mathbf{y})$  por la variable  $\tilde{\mathbf{x}}$ ,

$$\frac{\partial r_g(\mathbf{y})}{\partial \tilde{\mathbf{x}}} = 2 \int \mathbf{B}(\tilde{\mathbf{y}} - \mathbf{x}) p(\mathbf{x}|\mathbf{y}) d\mathbf{x} = 0 \quad (2)$$

La igualdad (2) se cumple cuando  $\tilde{\mathbf{x}} \int p(\mathbf{x}|\mathbf{y}) d\mathbf{x} = \int \mathbf{x} p(\mathbf{x}|\mathbf{y}) d\mathbf{x}$ .

Asumiendo que se cumple la condición de normalización para  $p(\mathbf{x}|\mathbf{y})$ , entonces se tiene

$$\tilde{\mathbf{x}} = \int \mathbf{x} p(\mathbf{x}|\mathbf{y}) d\mathbf{x} \quad (3)$$

Por lo tanto, la estimación con mínimo error cuadrático medio corresponde al valor de la FDP a posteriori  $p(\mathbf{x}|\mathbf{y})$ . De (3), se observa que la estimación óptima no depende de la matriz  $\mathbf{B}$ , lo que significa se tiene un óptimo general por todas las componentes del proceso  $\mathbf{x}$ .

**Ejemplo 2.26.** En el ejemplo anterior, suponer que se tiene la función simple de pérdida, definida en (2.80c).

En este caso, la minimización del riesgo a posteriori (2.90), que toma la forma

$$r_g(\mathbf{y}) = \frac{1}{\varepsilon} \int p(\mathbf{x}|\mathbf{y}) d\mathbf{x}$$

corresponde a la selección del valor  $\tilde{\mathbf{x}}$ , que maximiza la expresión

$$\frac{1}{\varepsilon} - r_g(\mathbf{y}) = \frac{1}{\varepsilon} - \frac{1}{\varepsilon} \int p(\mathbf{x}|\mathbf{y}) d\mathbf{x} = \frac{1}{\varepsilon} \int p(\mathbf{x}|\mathbf{y}) d\mathbf{x}$$

donde la integración se realiza sobre el espacio  $\mathfrak{E}_1$ , en el cual  $\|\epsilon\|_{\mathbf{B}} \geq \epsilon/2$ , o sobre el espacio  $\mathfrak{E}_0$ , donde se cumple  $\|\epsilon\|_{\mathbf{B}} < \epsilon/2$ . El empleo de la función simple de pérdidas implica un valor relativamente pequeño para el intervalo  $\epsilon$  ( $\epsilon \rightarrow 0$ ). Entonces, para obtener el máximo de la última expresión como valor de la estimación óptima  $\tilde{\mathbf{x}}$ , en cada trayectoria  $\mathbf{y}$  dada, se debe seleccionar el valor de  $\mathbf{x}$ , para el cual  $p(\mathbf{x}|\mathbf{y})$  toma el máximo valor posible, esto es, la estimación  $\tilde{\mathbf{x}}$  corresponde a la moda  $p(\mathbf{x}|\mathbf{y})$ . Esta estimación se denomina *estimación por máximo de densidad de probabilidad a posteriori*.

Conocida la trayectoria  $\mathbf{y}$ , la estimación óptima  $\tilde{\mathbf{x}}$  se halla como la raíz de la ecuación

$$\left. \frac{\partial p(\mathbf{x}|\mathbf{y})}{\partial \mathbf{x}} \right|_{\mathbf{x}=\tilde{\mathbf{x}}} = 0 \quad (4)$$

El valor de la estimación óptima  $\tilde{\mathbf{x}}$  se puede obtener mediante la maximización de la función  $p(\mathbf{x}|\mathbf{y})$  con relación a  $\mathbf{x}$ . Basados en el teorema de Bayes, (2.4), se tiene la relación condicional  $p(\mathbf{x}|\mathbf{y}) = p(\mathbf{x}|\mathbf{y})p(\mathbf{x})/p(\mathbf{y})$ , esto es,  $\ln p(\mathbf{x}|\mathbf{y}) = \ln p(\mathbf{x}, \mathbf{y}) - \ln p(\mathbf{y}) = \ln p(\mathbf{x}|\mathbf{y}) + \ln p(\mathbf{x}) - \ln p(\mathbf{y})$ .

Debido que  $p(\mathbf{y})$  no depende la variable  $\mathbf{x}$ , la maximización de  $p(\mathbf{x}|\mathbf{y})$  se reduce a hallar el valor máximo de la densidad de probabilidad conjunta  $p(\mathbf{x}, \mathbf{y})$ . De esta manera, el valor óptimo de la estimación se puede hallar como la raíz de una de las siguientes ecuaciones:

$$\left. \frac{\partial p(\mathbf{x}, \mathbf{y})}{\partial \mathbf{x}} \right|_{\mathbf{x}=\tilde{\mathbf{x}}} = \left. \frac{\partial p(\mathbf{y}|\mathbf{x})p(\mathbf{x})}{\partial \mathbf{x}} \right|_{\mathbf{x}=\tilde{\mathbf{x}}} = 0 \quad (5a)$$

$$\left. \frac{\partial \ln p(\mathbf{x}, \mathbf{y})}{\partial \mathbf{x}} \right|_{\mathbf{x}=\tilde{\mathbf{x}}} = \left( \left. \frac{\partial \ln p(\mathbf{y}|\mathbf{x})}{\partial \mathbf{x}} + \frac{\partial \ln p(\mathbf{x})}{\partial \mathbf{x}} \right) \right|_{\mathbf{x}=\tilde{\mathbf{x}}} = 0 \quad (5b)$$

La solución de la función de pérdida, en tareas de aplicación concretas, en parte, se realiza atendiendo al principio de implementación simple de los algoritmos óptimos de estimación. Desde este punto de vista, como se observa de (3) y (4), la estimación óptima, para el caso de la función simple de pérdida, se determina por las propiedades locales de FDP a posteriori de las componentes de estimación del vector  $\mathbf{x}$ , en la cercanía de su máximo global, en cambio, cuando se tiene la función cuadrática de pérdida, la estimación se determina por la forma de cambio de la FDP a priori, pero en todo el intervalo del vector  $\mathbf{x}$ . Por esta razón, es de esperar que para algunas tareas de estimación sea preferible la realización del algoritmo (4), que exigen menor detalle de la descripción de la FDP a posteriori, que la requerida para (3), aunque es posible que su rendimiento sea peor.

### 2.3.3. Decisión no Bayesiana

La aproximación Bayesiana implica que sobre el proceso  $\mathbf{x}$ , siendo aleatorio, se conoce su medida probabilística, en particular, su FDP. En caso contrario, se emplean diferentes métodos no Bayesianos de selección de la mejor función de decisión.

**Método de máxima verosimilitud.** Se considera que las mejores estimaciones bayesianas se obtienen por el método del máximo de probabilidad a posteriori, cuando la estimación se calcula mediante la maximización de la FDP conjunta  $p(\mathbf{x}, \mathbf{y}) = p(\mathbf{y}|\mathbf{x})p(\mathbf{x})$ . Sin embargo, en la práctica es usual que no se tenga información a priori alguna sobre el proceso  $\mathbf{x}$ , caso en el cual, es imposible obtener la solución de las ecuaciones (5a) y (5b) del ejemplo 2.26. Por lo tanto, una forma aproximada de la estimación óptima  $\tilde{\mathbf{x}}$  se halla simplificando el sistema de ecuaciones que contengan solo la función de verosimilitud. En este caso, la

estimación obtenida se denomina de *máxima verosimilitud* y se determina por la solución de una de las siguientes ecuaciones,

$$\left. \frac{\partial p(\mathbf{y}|\mathbf{x})}{\partial \mathbf{x}} \right|_{\mathbf{x}=\tilde{\mathbf{x}}} = 0 \quad (2.91a)$$

$$\left. \frac{\partial \ln p(\mathbf{y}|\mathbf{x})}{\partial \mathbf{x}} \right|_{\mathbf{x}=\tilde{\mathbf{x}}} = 0 \quad (2.91b)$$

El método de máxima verosimilitud se puede analizar como un caso particular de la estimación Bayesiana. Así por ejemplo, si en alguna de las expresiones (2.91a) ó (2.91b), se asume que el proceso aleatorio  $\mathbf{x}$  tiene FDP Gaussiana y para un aumento ilimitado de las varianzas de todas las componentes del vector  $\mathbf{x}$ , cuando se cumple que  $\partial \ln p(\mathbf{x})/\partial \mathbf{x} = 0$ , entonces la expresión (5b) se transforma en (2.91b). A propósito, al mismo resultado se llega asumiendo que la FDP es uniforme.

En general, el cambio de (5a) a (2.91a), o bien de (5b) a (2.91b) es aceptable y está condicionado al hecho de que la observación,  $\mathbf{y}$ , contenga tanta información sobre el verdadero valor de  $\mathbf{x}$ , que su estimación  $\tilde{\mathbf{x}}$  tenga de forma significativa una menor dispersión con respecto al valor verdadero  $\mathbf{x}$ , la cual debe ser menor que la dispersión determinada de forma a priori por la densidad  $p(\mathbf{x})$ . Además, se debe tener seguridad que  $p(\mathbf{x})$  no contiene picos muy altos en las regiones de definición de  $\mathbf{x}$ . En algunos casos, cuando la información de la FDP a priori es insuficiente, se asumen aproximaciones Gaussianas o uniformes y se obtienen estimaciones denominadas *seudobayesianas* [12].

El error cuadrático medio de estimación, para el método de máxima verosimilitud, es mayor que para el caso de estimación pseudobayesiana, y por su puesto mucho mayor que para el caso Bayesiano. La razón de este comportamiento del error de estimación está en que la construcción del estimador Bayesiano incluye la información a priori del vector estimado,  $\mathbf{x}$ , mientras en el caso de máxima verosimilitud, ésta se ignora o bien se toma igual para el peor caso.

**Método minimax.** De la expresión (2.86), se observa que a cada vector  $\mathbf{x}$  le corresponde un valor de riesgo condicional  $r_g(\mathbf{x})$ , lo que significa que este valor cambia para cada una de las señales, y por lo tanto, existirá un valor máximo  $r_g(\mathbf{x})_{\max}$  para todos los posibles valores de  $\mathbf{x}$ .

La función de decisión  $\hat{g}(\mathbf{y})$  determina la solución minimax, si se cumple que

$$\sup \hat{r}_g(\mathbf{x}) \leq \sup r_g(\mathbf{x})$$

Asumiendo la existencia de los valores extremos, la anterior relación toma la forma,

$$\min_g \max_{\mathbf{x}} r_g(\mathbf{x}) = \max_{\mathbf{x}} \hat{r}_g(\mathbf{x}) \quad (2.92)$$

El método de estimación minimax, descrito por (2.92), implica que se realiza la minimización por todo  $g$  del riesgo condicional  $r_g(\mathbf{x})$ , para el peor de los casos, por todos los

valores de  $\mathbf{x}$ . Además, la función de decisión  $\hat{g}(\mathbf{x})$  minimiza, por todos los  $\mathbf{x}$ , el máximo valor de la función de riesgo condicional  $r_g(\mathbf{x})$ . La principal particularidad del método (2.92) está en que garantiza un cierto valor de éxito, referido al peor de los casos a ser considerado. Sin embargo, la estimación puede resultar demasiado lejana de las situaciones con mayor probabilidad de ocurrencia.

La solución minimax en muchas tareas de aplicación real significa un procedimiento bastante complejo de resolver. Aunque si se asocia el método minimax al Bayesiano, la complejidad de solución disminuye. Así por ejemplo, la función de decisión minimax  $\hat{g}(\mathbf{x})$ , dadas ciertas restricciones de carácter muy suave (que son simples de cumplir), corresponde a la función Bayesiana con respecto a la distribución a priori menos favorable, que a propósito, frecuentemente resulta ser la FDP uniforme.

**Ejemplo 2.27.** Sea  $x(t)$  un proceso aleatorio escalar, cuyos valores deben localizarse dentro del intervalo de trabajo  $-a < x < a$ ,  $x \in \mathbb{R}$ , en caso contrario, el sistema se considera desintonizado y debe generarse una señal de alarma, por lo menos, mientras el proceso no regrese al intervalo indicado de trabajo. Hallar la estimación óptima  $\tilde{x}(t)$ , dada la trayectoria de observación  $y(t)$  con valores pertenecientes al intervalo  $(t_0, t)$ , que corresponde a una función conocida del proceso estimado mas una perturbación.

La respectiva decisión  $d = g(y)$  es el resultado de la filtración, esto es,  $g(y) = \tilde{x}(t)$ , que se determina por alguna de las funciones de pérdida, por ejemplo, la forma simple (2.80c),

$$f_c(e) = \begin{cases} 0, & |e| \leq \mu \\ c, & |e| > \mu \end{cases}$$

donde  $e(t) = x(t) - \tilde{x}(t)$ ,  $c = \text{const.}$  y  $\mu = \text{const.}$

El sistema puede generar error en dos situaciones: una, cuando la estimación está dentro del intervalo permitido, mientras el valor verdadero no lo está (error de primer tipo), o cuando la estimación está fuera del intervalo, pero el valor verdadero aún se conserva en el intervalo de sintonización (error de segundo tipo). En el primer caso, la función de pérdidas se asume igual  $A = \text{const.}$ , y para el segundo caso,  $B = \text{const.}$

## Problemas

**Problema 2.21.** Sea la trayectoria  $\mathbf{y} = \{y_k : k = 1, \dots, n\}$ ,  $n \rightarrow \infty$ , con distribución  $\mathcal{N}(\theta, 1)$ , sobre la cual se estima el parámetro  $\theta$ , para una función de riesgo  $f_c(\varepsilon) = (\theta - \varepsilon)^2$ . Calcular la función de pérdida para los siguientes casos de función de decisión Bayesiana:  $D_1 = m_{1\theta}$ ,  $D_2 = m_e(\theta)$ ,  $D_3 \equiv 0$ .

**Problema 2.22.** Resolver el problema anterior empleando los métodos de máxima verosimilitud y minimax para la selección de la mejor función de decisión.

**Problema 2.23.** Sea la observación  $\mathbf{y} = \{y_k : k = 1, \dots, n\}$  con FDP Gaussiana  $\mathcal{N}(\theta, \sqrt{\theta})$ ,  $\theta > 0$ . Estimar el parámetro  $\theta$  empleando la función de pérdida  $f_c(\varepsilon)$  dada en (2.80a) [20].

# Capítulo 3

## Representación de señales aleatorias

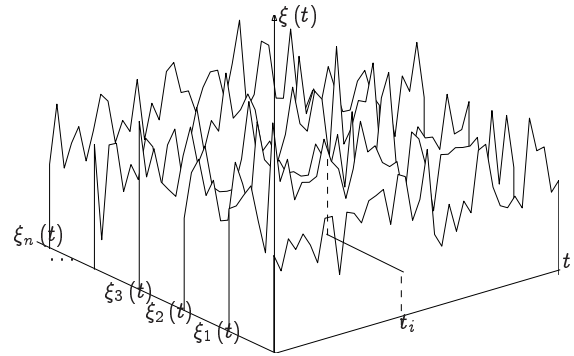
La *Teoría de procesos estocásticos* estudia las variables aleatorias, que dependen de un parámetro susceptible de cambiar de valor

B.R. Levin

La señal aleatoria  $\xi(s)$  corresponde a un proceso, que se desarrolla sobre la variable  $s$ . Cuando la variable del argumento de la señal aleatoria, que usualmente es el tiempo, discurre continuamente, entonces se habla de *funciones aleatorias*. Si en cambio, la variable del tiempo corresponde a una malla de valores discretos,  $s_1, s_2, \dots, s_n$ , entonces se habla de sucesiones o *series aleatorias*. El análisis de las señales aleatorias, en gran medida, depende de si cambia o no su estructura de aleatoriedad con respecto a su variable del argumento.

### 3.1. Señales aleatorias en el tiempo

Los fenómenos, que se desarrollan a lo largo del argumento correspondiente al tiempo y descritos por una señal aleatoria  $\xi(t)$ , se pueden analizar a partir de un conjunto longitudinal de mediciones o registros de valores  $\{\xi_i(t) \in \xi : i = 1, \dots, N\}$  como se ilustra en la Figura 3.1. En este caso, se define como *trayectoria* u observación a cada una de las mediciones simultáneas  $\xi_i$  de un misma señal aleatoria. Se define como *ensamble* o conjunto a todas las posibles trayectorias,  $\xi_i(t)$ , medidas o registradas en un intervalo de observación, que se relacionan con un mismo fenómeno aleatorio,  $\xi(t)$ .



**Figura 3.1.** Trayectorias de una señal aleatoria

En cualquier caso, las señales aleatorias pueden ser descritas mediante un ensamble de múltiples observaciones, entonces, se generan dos clases diferentes de promedios: se pueden efectuar mediciones sucesivas a lo largo de una misma observación  $\xi(t)$  a partir de las cuales se hallan sus momentos y valores de aleatoriedad o *valores promedios de tiempo*  $E\{\xi_i^n(t)\}$ , así mismo, se pueden examinar todas las observaciones del ensamble, en un momento de tiempo dado  $t_i$  con lo cual se hallan los *valores promedios de ensamble*  $E\{\xi^n(t_i)\}$ .

### 3.1.1. Estacionariedad de las señales aleatorias

Las características básicas de la estructura de aleatoriedad de las señales corresponden a los momentos y valores de aleatoriedad descritos en el numeral §2.1, los cuales pueden cambiar en el tiempo. Así por ejemplo, la FDP con dimensión múltiple  $p(\xi_1, \xi_2, \dots, \xi_n; t_1, t_2, \dots, t_n)$  corresponde a los valores instantáneos de las respectivas FDP singulares  $\xi_n(t_n)$ .

Un proceso se considera *estacionario* cuando su estructura de aleatoriedad no cambia en el tiempo, en particular, un proceso estocástico se define como *estacionario en el sentido angosto*, si sus funciones de probabilidad no son variables en el tiempo, esto es:

$$p(\xi_i, t) = p(\xi_i, t + \Delta t), \quad \forall \Delta t$$

condición, que prácticamente es bastante difícil de comprobar.

En cambio, un proceso estocástico  $\xi(t)$  se define como *estacionario en el sentido amplio*, cuando todos sus momentos y valores de aleatoriedad no varían para cualquiera que sean los tiempos de análisis  $(t_1, t_2, \dots)$ .

$$\begin{aligned} E\{\xi^n(t_n)\} &= \text{const.} \\ E\{(\xi(t_m) - m_{1\xi}(t_m))(\xi(t_n) - m_{1\xi}(t_n))\} &= \text{const} \end{aligned}$$

En consecuencia, los momentos (2.6), (2.9), (2.44) y (2.45) de las señales estacionarias se asocian con los siguientes *promedios de tiempo*:

- *Valores medios*

$$E\{\xi^n(t)\} = \int_{-\infty}^{\infty} \xi^n(t) dt \triangleq \overline{\xi^n(t)}, \quad n \in \mathbb{N} \quad (3.1)$$

- *Valores medios centralizados*

$$E\{\xi^n(t) - \overline{\xi(t)}\} = \int_{-\infty}^{\infty} (\xi(t) - \overline{\xi(t)})^n dt, \quad n \geq 2, n \in \mathbb{N} \quad (3.2)$$



- Valores de correlación

$$\begin{aligned}
 K_{\xi\eta}(t_m, t_n) &= E\{(\xi(t_m) - m_{1\xi}(t_m))(\eta(t_n) - m_{1\eta}(t_n))\} \\
 &= \int_{-\infty}^{\infty} (\xi(t_m) - m_{1\xi}(t_m))(\eta^*(t_n) - m_{1\eta}(t_n)) dt \\
 &= R_{\xi\eta}(t_m, t_n) - m_{1\xi}(t_m) m_{1\eta}(t_n)
 \end{aligned} \tag{3.3a}$$

$$R_{\xi\eta}(t_m, t_n) = E\{\xi(t_m)\eta^*(t_n)\} = \int_{-\infty}^{\infty} \xi(t_m)\eta^*(t_n) dt \tag{3.3b}$$

Cuando en los núcleos de valores de correlación se tiene el caso de dos variables aleatorias diferentes,  $\xi(t) \neq \eta(t)$ , se habla de la *función de correlación mutua*, mientras en el caso de análisis de una misma función,  $\xi(t) = \eta(t)$ , se dice de *función de correlación propia*. Si se tienen dos variables aleatorias  $\xi(t)$  y  $\eta(t)$  con los respectivos valores de correlación propia  $R_{\xi}(t_m, t_n)$  y  $R_{\eta}(t_m, t_n)$ , en analogía con la matriz (2.54), de forma adicional se considera la *matriz de correlación*:

$$\mathbf{R} = \begin{bmatrix} R_{\xi}(t_m, t_n) & R_{\xi\eta}(t_m, t_n) \\ R_{\eta\xi}(t_m, t_n) & R_{\eta}(t_m, t_n) \end{bmatrix}$$

**Ejemplo 3.1.** Sea la suma de dos señales aleatorias

$$\eta(t) = \xi_1(t) + \xi_2(t)$$

El valor medio del proceso resultante se determina como:

$$\begin{aligned}
 E\{\eta(t)\} &= E\{\xi_1(t) + \xi_2(t)\} = E\{\xi_1(t)\} + E\{\xi_2(t)\} \\
 &= m_{1\xi_1}(t) + m_{1\xi_2}(t)
 \end{aligned}$$

Mientras, la función de correlación de la señal aleatoria resultante será:

$$R_{\eta}(t_m, t_n) = E\{\xi_1(t_m) - m_{1\xi_1}(t_m)\} E\{\xi_2(t_n) - m_{1\xi_2}(t_n)\}$$

que se calcula como

$$\begin{aligned}
 &= E\{\xi_1(t_m) - m_{1\xi_1}(t_m) + \xi_2(t_m) - m_{1\xi_2}(t_m)\} E\{\xi_1(t_n) - m_{1\xi_1}(t_n) + \xi_2(t_n) - m_{1\xi_2}(t_n)\} \\
 &= E\{\xi_1(t_m) - m_{1\xi_1}(t_m)\} E\{\xi_1(t_n) - m_{1\xi_1}(t_n)\} + E\{\xi_1(t_m) - m_{1\xi_1}(t_m)\} \\
 &\quad \times E\{\xi_2(t_n) - m_{1\xi_2}(t_n)\} + E\{\xi_2(t_m) - m_{1\xi_2}(t_m)\} E\{\xi_1(t_n) - m_{1\xi_1}(t_n)\} \\
 &\quad + E\{\xi_2(t_m) - m_{1\xi_2}(t_m)\} E\{\xi_2(t_n) - m_{1\xi_2}(t_n)\}
 \end{aligned}$$

con lo cual se tiene que

$$R_{\eta}(t_m, t_n) = R_{\xi_1}(t_m, t_n) + R_{\xi_1\xi_2}(t_m, t_n) + R_{\xi_2\xi_1}(t_m, t_n) + R_{\xi_2}(t_m, t_n)$$

Al considerar que las variables  $\xi_1(t)$  y  $\xi_2(t)$  son no correlacionadas, entonces

$$R_{\eta}(t_m, t_n) = R_{\xi_1}(t_m, t_n) + R_{\xi_2}(t_m, t_n)$$

Asumiendo que se analiza la suma de la señal aleatoria  $\xi_1(t)$  con otra señal no aleatoria  $x(t)$ , se obtiene el siguiente valor medio:

$$\begin{aligned} E\{\eta(t)\} &= E\{\xi_1(t) + x(t)\} \\ &= E\{\xi_1(t)\} + x(t) \end{aligned}$$

con la respectiva función de correlación

$$R_\eta(t_m, t_n) = R_{\xi_1}(t_m, t_n)$$

Si la variable  $x(t)$  se convierte en constante  $x(t) = c$ , se obtienen los momentos:

$$\begin{aligned} E\{\eta(t)\} &= E\{\xi_1(t) + c\} = m_{\xi_1}(t) + m_{1c} \\ R_\eta(t_m, t_n) &= R_{\xi_1}(t_m, t_n) + \sigma_c^2 \end{aligned}$$

Se puede demostrar, que para la suma  $\eta(t) = a\xi_1(t) + b\xi_2(t)$  se tienen los momentos:

$$\begin{aligned} E\{\eta(t)\} &= am_{1\xi_1}(t) + bm_{1\xi_2}(t) \\ R_\eta(t_m, t_n) &= a^2 R_{\xi_1}(t_m, t_n) + b^2 R_{\xi_2}(t_m, t_n) + ab(R_{\xi_1\xi_2}(t_m, t_n) + R_{\xi_2\xi_1}(t_m, t_n)) \end{aligned}$$

**Ejemplo 3.2.** Hallar la función de correlación de la siguiente suma de señales aleatorias

$$\xi(t) = \sum_{n=1}^N \xi_n(t)$$

donde  $\xi_n(t) = \xi_{nc} \cos \omega_n t + \xi_{ns} \sin \omega_n t$ , siendo  $\xi_c$  y  $\xi_s$  valores aleatorios no correlacionados con media cero e igual varianza  $\sigma_{\xi_{nc}}^2 = \sigma_{\xi_{ns}}^2 = \sigma_{\xi_n}^2$ .

La función de correlación de  $\xi_n(t) = \xi_{nc} \cos \omega_n t + \xi_{ns} \sin \omega_n t$  se calcula como:

$$\begin{aligned} R_{\xi_n}(t_1, t_2) &= E\{\xi_n(t_1)\xi_n(t_2)\} \\ &= E\{(\xi_{nc} \cos \omega_n t_1 + \xi_{ns} \sin \omega_n t_1)(\xi_{nc} \cos \omega_n t_2 + \xi_{ns} \sin \omega_n t_2)\} \\ &= \sigma_{\xi_{nc}}^2 \cos \omega_n t_1 \cos \omega_n t_2 + \sigma_{\xi_{ns}}^2 \sin \omega_n t_1 \sin \omega_n t_2 \\ &= \sigma_{\xi_n}^2 \cos \omega_n (t_1 - t_2) \end{aligned}$$

que da como resultado

$$R_\xi(t_1, t_2) = \sum_{n=1}^N R_{\xi_n}(t_1, t_2) = \sum_{n=1}^N \sigma_{\xi_n}^2 \cos \omega_n (t_1 - t_2)$$

**Transformación lineal de señales aleatorias.** Sea un sistema con entrada  $x$  y salida  $y$ , descrito por un operador lineal  $\mathcal{K}$ , tal que cumpla las siguientes condiciones:

1.  $y = \mathcal{K}\{\lambda x\} = \lambda \mathcal{K}\{x\}$ ,  $\lambda = \text{const.}$
2.  $y = \mathcal{K}\{x_1 + x_2\} = \mathcal{K}\{x_2\} + \mathcal{K}\{x_1\}$

En los casos de transformación de variables, son importantes los valores de aleatoriedad y momentos de los procesos de entrada o salida, dada la descripción del sistema en la forma  $\eta = \mathcal{K}\{\xi\}$ .

En particular, el valor medio y la función de correlación de la salida son de la forma:

$$m_{1\eta}(t) = E\{\eta(t)\} = E\{\mathcal{K}\{\xi(t)\}\} = \mathcal{K}\{E\{\xi(t)\}\} \quad (3.4a)$$

$$\begin{aligned} R_{\eta}(t_m, t_n) &= E\{\eta(t_m) - m_{1\eta}(t_m)\} E\{\eta(t_n) - m_{1\eta}(t_n)\} \\ &= E\{\mathcal{K}\{\xi(t_m) - m_{1\xi}(t_m)\}\} E\{\mathcal{K}\{\xi(t_n) - m_{1\xi}(t_n)\}\} \\ &= \mathcal{K}\{E\{\xi(t_m) - m_{1\xi}(t_m)\}\} \mathcal{K}\{E\{\xi(t_n) - m_{1\xi}(t_n)\}\} \\ &= \mathcal{K}_{(t_m)} \mathcal{K}_{(t_n)} \{E\{\xi(t_m) - m_{1\xi}(t_m)\} E\{\xi(t_n) - m_{1\xi}(t_n)\}\} \\ &= \mathcal{K}_{(t_m)} \mathcal{K}_{(t_n)} \{R_{\xi}(t_m, t_n)\} \end{aligned} \quad (3.4b)$$

Cabe anotar que la suposición de estacionariedad simplifica las expresiones (3.4a) y (3.4b), obtenidas para la transformación de señales, así:

$$E\{\eta(t)\} = E\{\mathcal{K}\{\xi(t)\}\} = \mathcal{K}\{E\{\xi(t)\}\} = m_{1\eta} \quad (3.5a)$$

$$\mathcal{K}_{(t_m)} \mathcal{K}_{(t_n)} \{R_{\xi}(t_m, t_n)\} = \mathcal{K} \mathcal{K}_{(\tau)} \{R_{\xi}(\tau)\}, \quad \tau = t_m - t_n \quad (3.5b)$$

De otra parte, la reacción de un sistema lineal e invariante en el tiempo (1.39), de acuerdo con (1.33), se describe por la reacción del operador lineal del sistema a la función  $\delta(t)$ , o respuesta a impulso  $h(t)$ , por lo que si a la entrada del sistema se tiene la señal aleatoria  $\xi(t)$ , mientras a la salida se tiene la señal aleatoria  $\eta(t)$ , entonces, se observan las siguientes relaciones en el tiempo entre los momentos de entrada y salida del sistema:

$$m_{1\eta}(t) = \int_0^{\infty} m_{1\xi}(t - \tau) h(\tau) d\tau \quad (3.6a)$$

$$R_{\eta}(t_m, t_n) = \int_0^{\infty} \int_0^{\infty} R_{\xi}(t_m - \tau_m, t_n - \tau_n) h(\tau_m) h(\tau_n) d\tau_m d\tau_n \quad (3.6b)$$

**Ejemplo 3.3.** Dados, para la variable  $\xi(t)$ , la media  $m_{1\xi}(t) = at^2$  y la función de correlación

$$R_{\xi}(t_m, t_n) = \sigma_{\xi}^2 \exp -\alpha |t_m - t_n|$$

Hallar los momentos  $m_{1\eta}$ ,  $R_{\eta}(t_m, t_n)$  y varianza de la señal aleatoria en los siguientes dos casos de análisis de transformación lineal:

$$a) \eta(t) = \frac{d\xi}{dt}, \quad b) \eta(t) = \int_0^t \xi(t) dt$$

En el caso a), el valor medio es igual a:

$$\begin{aligned} m_{1\eta}(t) &= \mathcal{K}\{E\{\xi(t)\}\} = \frac{d}{dt} m_{1\xi}(t) \\ &= 2at \end{aligned}$$

La función de correlación del proceso de salida se halla diferenciando dos veces la función de

correlación del proceso de entrada, de la siguiente manera:

$$\begin{aligned} t_m > t_n : R_\xi(t_m, t_n) &= \sigma_\xi^2 e^{-\alpha(t_m - t_n)} \\ \frac{\partial}{\partial t_m} R_\xi(t_m, t_n) &= -\sigma_\xi^2 \alpha e^{-\alpha(t_m - t_n)} \\ \frac{\partial^2}{\partial t_m \partial t_n} R_\xi(t_m, t_n) &= \sigma_\xi^2 \alpha^2 e^{-\alpha(t_m - t_n)} \end{aligned}$$

$$\begin{aligned} t_m < t_n : R_\xi(t_m, t_n) &= \sigma_\xi^2 e^{-\alpha(t_n - t_m)} \\ \frac{\partial}{\partial t_m} R_\xi(t_m, t_n) &= \sigma_\xi^2 \alpha e^{-\alpha(t_n - t_m)} \\ \frac{\partial^2}{\partial t_m \partial t_n} R_\xi(t_m, t_n) &= -\sigma_\xi^2 \alpha^2 e^{-\alpha(t_n - t_m)} \end{aligned}$$

Juntar ambos casos de análisis ( $t_m > t_n$  y  $t_m < t_n$ ) es imposible. Por lo tanto, se puede inferir que la función de correlación  $R_\xi(t_m, t_n)$  no tiene la segunda derivada para el valor  $t_m = t_n$ , luego, el proceso  $\xi(t)$  no es diferenciable.

En el caso b), la media  $m_{1\xi}(t) = at^3/3$ . Mientras, la función de correlación se define como

$$R_\xi(t_m, t_n) = \sigma_\xi^2 \int_0^{t_m} \int_0^{t_n} e^{-\alpha|\lambda_m - \lambda_n|} d\lambda_m d\lambda_n$$

Luego, se tienen los siguientes casos de análisis:

$$\begin{aligned} R_\xi(t_m, t_n) &= \sigma_\xi^2 \int_0^{t_m} \int_0^{t_n} e^{-\alpha|\lambda_m - \lambda_n|} d\lambda_n d\lambda_m \\ t_n > t_m : \int_0^{t_n} e^{-\alpha|\lambda_m - \lambda_n|} d\lambda_n &= \int_0^{\lambda_m} e^{-\alpha(\lambda_m - \lambda_n)} d\lambda_n + \int_{\lambda_m}^{t_n} e^{-\alpha(\lambda_n - \lambda_m)} d\lambda_n \\ &= \frac{1}{\alpha} e^{-\alpha(\lambda_m - \lambda_n)} \Big|_0^{\lambda_m} - \frac{1}{\alpha} e^{\alpha(\lambda_m - \lambda_n)} \Big|_{\lambda_m}^{t_n} = \frac{1}{\alpha} \left( 2 - e^{-\alpha\lambda_m} - e^{-\alpha(t_n - \lambda_m)} \right); \\ &\int_0^{t_m} \frac{1}{\alpha} \left( 2 - e^{-\alpha\lambda_m} - e^{-\alpha(t_n - \lambda_m)} \right) d\lambda_m \\ &= \frac{1}{\alpha} \left( 2\lambda_m + \frac{1}{\alpha} e^{-\alpha\lambda_m} - \frac{1}{\alpha} e^{-\alpha(t_n - \lambda_m)} \right) \Big|_0^{t_m} \\ &= \frac{1}{\alpha^2} \left( 2\alpha t_m - 1 + e^{-\alpha t_m} + e^{-\alpha t_n} - e^{-\alpha(t_n - t_m)} \right) \\ R_\eta(t_m, t_n) &= \frac{\sigma_\xi^2}{\alpha^2} \left( 2\alpha t_m - 1 + e^{-\alpha t_m} + e^{-\alpha t_n} - e^{-\alpha(t_n - t_m)} \right) \end{aligned}$$

El análisis de  $t_m > t_n$ , muestra que de manera similar se obtiene:

$$R_\eta(t_m, t_n) = \frac{\sigma_\xi^2}{\alpha^2} \left( 2\alpha t_n - 1 + e^{-\alpha t_m} + e^{-\alpha t_n} - e^{-\alpha(t_m - t_n)} \right)$$

Al juntar ambas expresiones obtenidas en una sola función, que generalice para todo valor de  $t_m$  y  $t_n$ , se tiene:

$$R_\eta(t_m, t_n) = \frac{\sigma_\xi^2}{\alpha^2} \left( 2\alpha \min\{t_m, t_n\} - 1 + e^{-\alpha t_m} + e^{-\alpha t_n} - e^{-\alpha|t_m - t_n|} \right)$$

Finalmente, basados en lo anterior se puede encontrar la varianza, haciendo  $t_m = t_n = t$

$$\sigma_\eta^2(t) = 2\sigma_\xi^2 (\alpha t - 1 + e^{-\alpha t}) / \alpha^2$$

**Ejemplo 3.4.** Un proceso aleatorio corresponde a la transformación  $\xi(t) = \phi f(t)$ , siendo  $f$  una función determinística y  $\phi$  una variable aleatoria. Determinar si el proceso  $\xi(t)$  es estacionario.

La estacionariedad en el sentido amplio, exige la invariabilidad en el tiempo de los valores de aleatoriedad, por ejemplo del primer momento inicial (3.1) y la varianza (3.2),

$$\begin{aligned} E\{\xi(t)\} &= E\{\phi f(t)\} = f(t) E\{\phi\} \\ E\left\{\left(\xi(t) - \overline{\xi(t)}\right)^2\right\} &= E\left\{\left(\phi f(t) - \overline{\phi f(t)}\right)^2\right\} \\ &= f^2(t) E\{\phi\} \end{aligned}$$

que resultan variantes en el tiempo, y por ende, el proceso  $\xi(t)$  no es estacionario.

En la práctica, es frecuente la remoción del valor medio, y la consecuente normalización del proceso aleatorio  $\xi(t)$  sobre la desviación estándar,

$$\xi_s(t) = \frac{\xi(t) - \overline{\xi(t)}}{\sigma_\xi}$$

Sin embargo, hay que tener en cuenta que la FDP del proceso estandarizado es diferente de la densidad de probabilidad inicial, en particular, teniendo en cuenta la expresión para la transformación de variables (2.34), se obtiene

$$p(\xi) = \frac{1}{\sigma_\xi} p\left(\sigma_\xi \xi_s + \overline{\xi(t)}\right)$$

### 3.1.2. Ergodicidad de las señales aleatorias

Sea el proceso estacionario, en el cual se conocen los valores del primer momento inicial de cada una de las  $n$  trayectorias. Considerando la estacionariedad del proceso, la esperanza matemática del mismo en cualquier corte de tiempo  $t_i$ , como se muestra en la Figura 3.1, se determina de la forma

$$m_{1\xi(t_i)} = \frac{1}{n} \sum_{k=1}^n \xi_k(t_i)$$

El cálculo de primer momento, por alguna de las observaciones definida en un intervalo de análisis  $T$ , se puede realizar segmentando el intervalo en una malla de subintervalos,

determinados sobre los momentos de tiempo  $t_k = k\Delta t$ ,  $k = 1, \dots, n$ , esto es,  $n = T/\Delta t$ . La estimación de la esperanza sobre la malla de subintervalos se determina de la expresión:

$$\tilde{m}_{1\xi_i} = \frac{1}{n} \sum_{k=1}^n \xi_i(t_k) \Delta t$$

que al hallar el límite, resulta en

$$\tilde{m}_{1\xi_i} \underset{n \rightarrow \infty}{\approx} \frac{1}{T} \int_0^T \xi_i(t) dt$$

Por cuanto, se asume que las propiedades de aleatoriedad de los procesos estacionarios se mantienen invariantes en el tiempo, entonces, cuando se tiene un número  $n$  suficientemente grande, para cada trayectoria se cumple la igualdad

$$\tilde{m}_{1\xi_i} = m_{1\xi(t_i)}, \quad i = 1, \dots, n$$

Generalizando la expresión anterior, un proceso aleatorio estacionario se define como *ergódico*, cuando los valores promedios de tiempo y de ensamble son idénticos, esto es,

$$E \{ \xi_i^n(t) \} = E \{ \xi^n(t_i) \} \quad (3.7)$$

La media de cada observación del ensamble es igual a la media de la observación media del ensamble. De (3.7), se entiende que un proceso ergódico es estacionario, pero no viceversa.

La estimación de los momentos (3.1) para una señal ergódica  $\xi(t)$  se describe como

$$\tilde{E} \{ \xi^n(t) \} = \int_{-\infty}^{\infty} \xi^n(t) w(t) dt, \quad n \in \mathbb{Z} \quad (3.8)$$

siendo  $w(t)$  la función de peso, que cumple la restricción

$$\lim_{T \rightarrow \infty} \int_0^T w(t) dt = 1 \quad (3.9)$$

La función de peso rectangular es la más frecuentemente empleada en la estimación de los momentos en (3.8),

$$w(t) = \begin{cases} 1/T, & 0 \leq t \leq T \\ 0, & \text{otros valores de } t \end{cases} \quad (3.10)$$

La función de peso (3.10) corresponde a la respuesta a impulso (1.35) de un filtro lineal,

cuya respectiva función de transferencia es denominada *característica espectral*.

$$\Phi(\omega) = \mathcal{F}\{w(t)\} = \int_{-\infty}^{\infty} w(t) e^{-j\omega t} dt$$

que en el caso particular de (3.10) es igual a

$$\Phi(\omega) = e^{-j\omega T/2} \text{sinc}(\omega T/2)$$

Los valores de aleatoriedad dados en (3.1), (3.2), (3.3a) y (3.3b) para los procesos ergódicos, teniendo en cuenta la estimación (3.10), toman las respectivas expresiones:

$$E\{\xi^n(t)\} = \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T \xi^n(t) dt, \quad n \in \mathbb{N} \quad (3.11a)$$

$$E\{\xi^n(t) - \overline{\xi(t)}\} = \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T (\xi(t) - \overline{\xi(t)})^n dt, \quad n \geq 2, n \in \mathbb{N} \quad (3.11b)$$

$$R_\xi(\tau) = \lim_{T \rightarrow \infty} \frac{1}{T} \int_T (\xi(t))(\xi^*(t+\tau)) dt \quad (3.11c)$$

$$\begin{aligned} K_\xi(\tau) &= \lim_{T \rightarrow \infty} \frac{1}{T} \int_T (\xi(t) - \overline{\xi(t)}) (\xi^*(t+\tau) - \overline{\xi(t)}) dt \\ &= R_\xi(\tau) - m_{1\xi}^2 \end{aligned} \quad (3.11d)$$

Los momentos de las señales ergódicas se relacionan con el siguiente sentido físico:

- Valor medio  $\overline{\xi(t)}$  es la *componente constante*.
- El valor medio al cuadrado  $\overline{\xi(t)^2}$  corresponde a la *potencia de la componente directa*.
- Valor cuadrático medio  $\overline{\xi^2(t)}$  es la *potencia promedio*.
- La varianza  $\sigma_\xi^2$  es la *potencia de la componente alterna*.
- La desviación estándar  $\sigma_\xi$  corresponde al *valor rms*.
- La función de correlación propia corresponde a la *potencia de la componente alterna*

**Ejemplo 3.5.** Hallar la función de correlación de la señal aleatoria

$$\xi(t) = a \cos(\omega t + \phi)$$

siendo  $\phi$  la variable aleatoria.

A partir de la definición (3.11c) se tiene que

$$\begin{aligned}
R_{\xi}(\tau) &= \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T a \cos(\omega t + \phi) a \cos(\omega t + \omega \tau + \phi) dt \\
&= \lim_{T \rightarrow \infty} \frac{a^2}{T} \int_0^T \frac{1}{2} (\cos(\omega \tau + \phi) + \cos(2\omega t + \omega \tau + 2\phi)) dt
\end{aligned}$$

La primera integral se calcula de la forma

$$\frac{a^2}{2T} \int_0^T \cos \omega \tau dt = \frac{a^2}{2} \cos \omega \tau$$

Mientras, la segunda integral es igual a

$$\begin{aligned}
\frac{a^2}{2T} \int_0^T \cos(2\omega t + 2\phi + \omega \tau) dt \\
= \frac{a^2}{2T} \cos(2\phi + \omega \tau) \int_0^T \cos 2\omega t dt - \frac{a^2}{2T} \sin(2\phi + \omega \tau) \int_0^T \sin 2\omega t dt
\end{aligned}$$

Por cuanto,

$$\lim_{T \rightarrow \infty} \int_T \cos k\omega t dt = \lim_{T \rightarrow \infty} \text{sinc}(k\omega T) = 0, \quad \lim_{T \rightarrow \infty} \int_T \sin k\omega t dt = 0, \quad \forall k \in \mathbb{Z}$$

entonces, finalmente se obtiene la siguiente expresión para la función de correlación:

$$R_{\xi}(\tau) = \frac{a^2}{2} \cos \omega \tau$$

Por su papel preponderante en la descripción de procesos estacionarios que las funciones de correlación y covarianza tienen se debe tener en cuenta sus siguientes propiedades:

(a). *Paridad.*  $R_{\xi}(\tau) = R_{\xi}^*(-\tau)$ ,  $K_{\xi}(\tau) = K_{\xi}(-\tau)$

A partir de la estacionariedad del proceso se puede demostrar la simetría de los valores de correlación respecto a los argumentos  $t_1$  y  $t_2$ :

$$E \left\{ \xi_m(t_1) - \overline{\xi_m(t_1)}, \xi_n(t_2) - \overline{\xi_n(t_2)} \right\} = E \left\{ \xi_n(t_2) - \overline{\xi_n(t_2)}, \xi_m(t_1) - \overline{\xi_m(t_1)} \right\}$$

(b). *Valor máximo.*  $|R_{\xi}(\tau)| \leq R_{\xi}(0)$ ,  $|K_{\xi}(\tau)| \leq K_{\xi}(0)$ .

Por cuanto es evidente la siguiente desigualdad  $E \left\{ (\xi(t) \pm \xi(t + \tau))^2 \right\} \geq 0$ , abriendo paréntesis, se obtiene

$$\overline{\xi^2(t)} + \overline{\xi^2(t + \tau)} \pm 2\overline{\xi(t)\xi(t + \tau)} \geq 0$$



Debido a la estacionariedad del proceso, los dos primeros términos son iguales, con lo cual se obtiene la cota máxima de la función.

$$\overline{\xi^2(t)} \geq \pm \overline{\xi(t)\xi(t+\tau)} \Rightarrow R_\xi(0) \geq R_\xi(\tau)$$

Por cierto, de (3.3a) se tienen los siguiente valores en el origen:

$$K_\xi(0) = \frac{1}{T} \int_0^T \left( \xi(t) - \overline{\xi(t)} \right) \left( \xi^*(t) - \overline{\xi(t)} \right) dt = \sigma_\xi^2, \quad (3.12a)$$

$$R_\xi(0) = \frac{1}{T} \int_0^T \xi(t) \xi^*(t) dt = \overline{\xi^2(t)} \quad (3.12b)$$

- (c). *Continuidad.* Si la función  $R_\xi(\tau)$  es continua en  $\tau = 0$ , entonces  $R_\xi(\tau)$  también es continua para todo  $\tau$ .
- (d). *Periodicidad.*  $R_\xi(\tau) = R_\xi(\tau + T), \forall \tau \in T$ .
- (e). *Restricción en la forma.* Dado un proceso físicamente realizable, la transformada de Fourier de su función de correlación debe cumplir la siguiente restricción de forma:

$$\mathcal{F}\{R_\xi(\tau)\} \geq 0, \quad \forall \omega \quad (3.13)$$

- (f). *Convergencia.* Si el proceso aleatorio  $\xi(t)$  no es periódico, entonces se cumple,

$$\lim_{|\tau| \rightarrow \infty} R_\xi(\tau) = \overline{\xi^2(t)}, \quad \lim_{|\tau| \rightarrow \infty} K_\xi(\tau) = 0,$$

En este sentido se introduce el concepto de *intervalo de correlación*, que se puede determinar, bien como la parte  $\beta \ll 1$  del valor máximo de la función de correlación, bien como la mitad de la base del rectángulo  $\Delta\tau_e$  con altura unitaria y área igual a la de la función de correlación normalizada. En el primer caso, el intervalo de correlación corresponde al valor  $\tau = \tau_c$ , tal que el valor de la función normalizada de correlación sea considerablemente pequeño:

$$\frac{R_\xi(\tau_c)}{\sigma_\xi^2} = \beta \approx 0.05 \dots 0.1 \quad (3.14)$$

mientras en el segundo caso, el intervalo se calcula a partir de la integral

$$\Delta\tau_e = \frac{1}{R_\xi(0)} \int_0^\infty R_\xi(\tau) d\tau \quad (3.15)$$

Por último, se puede decir que la función de correlación caracteriza la dependencia estadística entre los valores instantáneos del proceso, dados en diferentes momentos del tiempo.

po. Así, por ejemplo, sea la diferencia cuadrática media entre dos valores instantáneos del proceso definida de la forma,

$$\varepsilon_{\xi}(\tau) = E \left\{ (\xi(t) - \xi(t + \tau))^2 \right\}$$

de la cual se observa que

$$\begin{aligned} \varepsilon_{\xi}(\tau) &= \overline{\xi^2(t)} + \overline{\xi^2(t + \tau)} - 2\overline{\xi(t)\xi(t + \tau)} = 2\overline{\xi^2(t)} - 2R_{\xi}(\tau) \\ &= 2R_{\xi}(0) - 2R_{\xi}(\tau) \end{aligned}$$

luego,  $2R_{\xi}(\tau) + \varepsilon_{\xi}(\tau) = 2R_{\xi}(0) = \text{const.}$

Por lo tanto, la función de correlación complementa la desviación cuadrática media hasta un valor constante y proporcional a la potencia del proceso.

**Ejemplo 3.6.** Hallar el intervalo de correlación para la función de correlación de la forma,  $R_{\xi}(\tau) = k/(1 - \alpha^2\tau^2)$ .

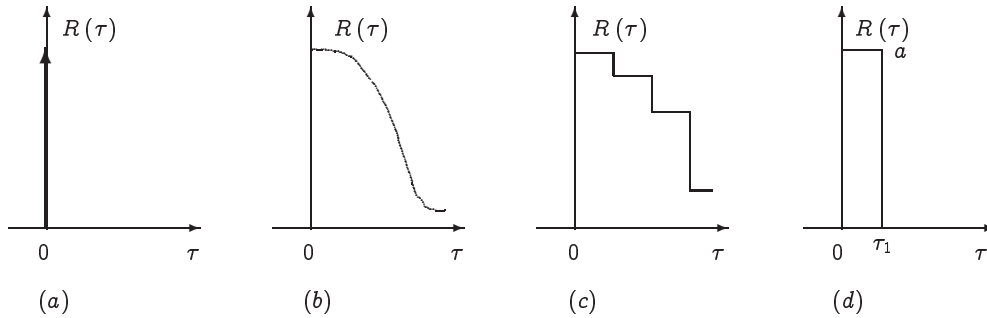
De la expresión (3.14), se tiene que

$$1/(1 - \alpha^2\tau_c^2) = \beta \Rightarrow \tau_c = \frac{1}{\alpha} \sqrt{\frac{1 - \beta}{\beta}}$$

mientras para el segundo caso (3.15) se obtiene

$$\Delta\tau = \int_0^{\infty} \frac{1}{1 + \alpha^2\tau^2} d\tau = \frac{1}{\alpha} \arctan \alpha\tau \Big|_0^{\infty} = \frac{\pi}{2\alpha}$$

Es usual considerar las siguientes formas de clasificación de los procesos ergódicos de acuerdo a su función de correlación propia:



**Figura 3.2.** Clasificación de funciones de correlación

- Procesos con correlación nula, en los cuales el valor de la función de correlación es cero, excepto para el corte de tiempo  $R(0) \sim \delta(\tau)$  (Figura 3.2(a)).

- Procesos con correlación continua, en los cuales el valor de la función cambia sobre un dominio continuo, (Figura 3.2(b)).
- Procesos con correlación discreta o *procesos de Markov*, cuando la función de correlación cambia de forma escalonada sobre una malla discreta de valores del intervalo de correlación  $\{\tau_k\}$ , (Figura 3.2(c)). Dentro de los procesos de Markov, una clase importante corresponde al caso cuando la función de correlación discreta se limita a un solo escalón (Figura 3.2(d)) lo que se interpreta como la dependencia de dos valores contiguos del proceso.

Realmente el proceso ilustrado en la Figura 3.2(d) es un modelo abstracto, en la medida en que la función de correlación no cumple con la condición de existencia (3.13), para todos los valores del intervalo de correlación  $\tau$ , como se observa de la relación:

$$\int_0^{\infty} R(\tau) \cos \omega \tau d\tau = \int_0^{\tau_1} a \cos \omega \tau d\tau = \frac{a}{\omega} \sin \omega \tau_1 \neq 0, \quad \forall \tau$$

### 3.1.3. Descomposición espectral de señales aleatorias

Del ejemplo 3.2, se observa que la función de correlación resultante  $R_{\xi}(t_1, t_2)$  es estacionaria, en la medida en que el argumento del coseno depende sólo de la distancia entre los valores de tiempo  $|t_2 - t_1|$ . Por lo que la función de correlación resultante es:

$$R_{\xi}(t_1, t_2) = R_{\xi}(\tau) = \sum_n \sigma_{\xi_n}^2 \cos \omega_n \tau, \quad \sigma_{\xi}^2 = \sum_n \sigma_{\xi_n}^2$$

Lo anterior, plantea la tarea de representación de señales en forma de la serie generalizada de Fourier (1.4), pero ampliada al caso de procesos aleatorios estacionarios.

Sea  $\xi(t)$  un proceso estacionario en el sentido amplio, dado en el intervalo  $(0, T)$ . Entonces, su descomposición ortogonal tendrá la forma:

$$\xi(t) = \sum_{n=0}^{\infty} \xi_n \phi_n(t) \tag{3.16}$$

donde los coeficientes de descomposición, en concordancia con (1.7), se determinan como:

$$\xi_n = \frac{1}{\mathcal{E}_n} \int_0^T \xi(t) \phi_n^*(t) dt, \quad \mathcal{E}_i = \int_0^T \phi_n^2(t) dt \tag{3.17}$$

Sin embargo, a diferencia de la representación (1.4), los coeficientes de (3.17) son aleatorios, por lo que la convergencia de la suma en (3.16) hacia la señal aleatoria  $\xi(t)$ , se debe entender en el sentido del valor cuadrático medio:

$$\lim_{N \rightarrow \infty} E \left\{ \left| \xi(t) - \sum_{n=0}^{\infty} \xi_n \phi_n(t) \right|^2 \right\} = 0$$

Uno de los problemas importantes a resolver es la selección del conjunto base de representación  $\{\phi_n(t)\}$ , el cual puede ser escogido por los dos siguientes principios: Primero, que se brinde el menor error de representación para un número  $N$  dado de coeficientes, o bien, segundo, que se genere el menor número  $N$  de coeficientes para un valor dado de error de representación. Así mismo, es importante que los coeficientes (3.17) tengan correlación cero entre ellos, por cuanto las tareas asociadas al proceso de señales aleatorias se resuelven de manera más fácil.

Se puede demostrar que para una señal aleatoria, representada por la serie (3.16), que tenga función de correlación continua  $R(t, \tau)$ , el valor de la esperanza del error cuadrático medio integral (potencia media de error), expresado como:

$$E \left\{ \int_0^T \left[ \xi(t) - \sum_{n=0}^{N-1} \xi_n \phi_n(t) \right]^2 dt \right\} \quad (3.18)$$

el error será el mínimo posible (3.18), para cualquier  $N$ , si el conjunto base de representación  $\{\phi_n(t) : n = 0, 1, \dots, N-1\}$  cumple la condición homogénea de Fredholm de segundo tipo [15, 22]:

$$\lambda_i \phi_i(t) = \frac{1}{T} \int_0^T R_\xi(t, \tau) g(\tau) d\tau \quad (3.19)$$

donde  $\{\phi_n(t)\}$  son las funciones propias (solución) de la ecuación y los coeficientes  $\lambda_n$  son los valores propios del núcleo  $R_\xi(t, \tau)$  de la ecuación. El conjunto de funciones propias  $\{\phi_n(t)\}$  es ortogonal, cuyos coeficientes pueden ser escogidos de tal manera que puedan ser ortonormales. Los coeficientes de descomposición (3.16) de la señal aleatoria, en caso de emplear el conjunto base de representación  $\{\phi_n(t)\}$  que cumplan con (3.19), resultan tener correlación nula. Además, si la señal aleatoria se asume del tipo Gaussiano, entonces los coeficientes resultan ser independientes estadísticamente. Así mismo, si se cumple que  $E\{\xi_n\} = 0$ , entonces, el valor de la varianza  $\sigma_{\xi_n}^2$  converge al respectivo valor propio  $\lambda_n$ .

Para el conjunto base de representación  $\{\phi_n(t)\}$ , que cumple la condición (3.19), la esperanza de la potencia de error (3.18), dada en el intervalo  $(0, T)$ , se determina como:

$$\begin{aligned} \bar{\varepsilon}_T^2 &= \frac{1}{T} E \left\{ \int_0^T \left\{ \xi(t) - \sum_{n=0}^{N-1} \xi_n \phi_n(t) \right\}^2 dt \right\} \\ &= \frac{1}{T} \int_0^T E \{ \xi^2(t) \} dt - \frac{2}{T} \int_0^T \sum_{n=0}^{N-1} E \{ \xi_n \xi^2(t) \} \phi_n(t) dt \\ &\quad + \frac{1}{T} \int_0^T \sum_{m=0}^{N-1} \sum_{n=0}^{N-1} E \{ \xi_m \xi_n \} \phi_m(t) \phi_n(t) dt \end{aligned}$$

Por lo que se tiene que

$$\bar{\varepsilon}_T^2 = \sigma_\xi^2 - \sum_{n=0}^{N-1} \sigma_{\xi_n}^2 \quad (3.20)$$

La expresión (3.20) permite hallar el número  $N$  de componentes de la serie de descomposición (3.16), que brinde un valor a priori dado de error de representación.

La descomposición de señales aleatorias, que tengan función de correlación continua, por la serie (3.16), en la cual el conjunto base de representación corresponde a las funciones propias, se denomina *descomposición de Karhunen-Loève* (K-L).

El ruido blanco Gaussiano, mostrado en el literal §3.1.4 con función de correlación (3.34), se puede descomponer en el intervalo  $(0, T)$  empleando cualquier conjunto base ortogonal, pero en todo caso, los coeficientes de descomposición serán valores aleatorios Gaussianos estadísticamente independientes con igual varianza de valor  $N_0/2$ .

**Ejemplo 3.7.** Sean  $\xi(t)$  y  $\eta(t)$ , señales aleatorias dadas en un intervalo finito de tiempo  $t \in T_\xi$ . Hallar la función de correlación mutua usando la descomposición ortogonal de Fourier. Las series de Fourier (1.8), para las señales aleatorias,  $\xi(t)$  y  $\eta(t)$ , tienen la misma forma,

$$x(t) = a_0 + \sum_{n=1}^{\infty} (a_n \cos n\Omega t + b_n \sin n\Omega t), \quad \Omega = \frac{2\pi}{T_\xi}, \quad x \in \{\xi, \eta\}$$

La descomposición de Fourier de la versión con desfase de cada proceso tiene la forma,

$$\begin{aligned} x(t + \tau) = & a_0 + \sum_{n=1}^{\infty} a_n \cos n\Omega \tau \cos n\Omega t - \sum_{n=1}^{\infty} a_n \sin n\Omega \tau \sin n\Omega t + \\ & + \sum_{n=1}^{\infty} b_n \sin n\Omega \tau \cos n\Omega t + \sum_{n=1}^{\infty} b_n \cos n\Omega \tau \sin n\Omega t \end{aligned}$$

Reemplazando ambas expresiones, en la definición de la función de correlación mutua,

$$R_{\xi\eta}(\tau) = a_{0\xi}a_{0\eta} + \frac{1}{2} \sum_{n=1}^{\infty} (a_{n\xi}a_{n\eta} + b_{n\xi}b_{n\eta}) \cos n\Omega \tau + \frac{1}{2} \sum_{n=1}^{\infty} (a_{n\xi}b_{n\eta} - a_{n\eta}b_{n\xi}) \sin n\Omega \tau$$

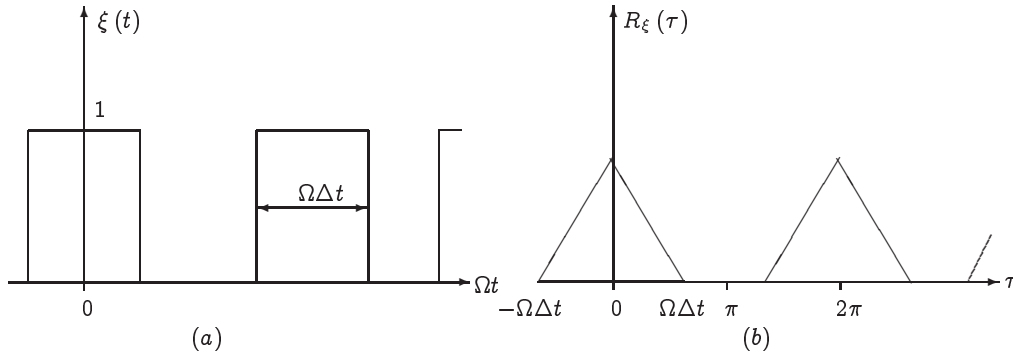
Cuando se asume la paridad de ambas señales aleatorias en análisis, entonces, los términos del senos se convierten en 0;  $b_\xi = b_\eta = 0$ , luego, la anterior expresión se simplifica hasta,

$$R_{\xi\eta}(\tau) = a_{0\xi}a_{0\eta} + \frac{1}{2} \sum_{n=1}^{\infty} a_{n\xi}a_{n\eta} \cos n\Omega \tau$$

**Ejemplo 3.8.** Hallar la función de correlación propia para el tren de pulsos cuadrados, mostrado en la Figura 3.3(a), empleando la descomposición ortogonal de Fourier.

La serie de Fourier de la función tren de pulsos cuadrados tiene la forma

$$\xi(t) = \frac{\Omega \Delta t}{2} + \frac{2}{\pi} \sum_{n=1}^{\infty} \frac{1}{n} \sin n \frac{\Omega \Delta t}{2} \cos n\Omega t$$



**Figura 3.3.** Función periódica pulso cuadrado

la cual se reemplaza por la expresión, obtenida en el ejemplo anterior para el caso de funciones pares, haciendo  $a_\xi = a_\eta$ , con lo que la función de correlación propia resulta en

$$\begin{aligned} R_\xi(\tau) &= a_{0\xi}^2 + \frac{1}{2} \sum_{n=1}^{\infty} a_{n\xi}^2 \cos n\Omega\tau \\ &= \left( \frac{\Omega\Delta t}{2\pi} \right)^2 + \frac{4}{\pi^2} \frac{1}{2} \sum_{n=1}^{\infty} \frac{1}{n^2} \sin^2 n \frac{\Omega\Delta t}{2} \cos n\Omega\tau \end{aligned}$$

Teniendo en cuenta la siguiente expresión cerrada, obtenida para la suma [13]:

$$\sum_{n=1}^{\infty} \sin^2 \frac{n\alpha}{n^2} \cos nx = \begin{cases} -\frac{\pi}{4}x + \frac{\pi\alpha}{2} - \frac{\alpha^2}{2}, & 0 < x < 2\alpha \\ -\frac{\alpha^2}{2}, & 2\alpha < x < \pi \end{cases}$$

entonces, se observa claramente, que la función de correlación corresponde a un tren de pulsos triangulares como se muestra en la Figura 3.3(b).

### 3.1.4. Densidad espectral de potencia

La descripción directa de una señal aleatoria  $\xi(t)$  en el dominio espectral, mediante la TF,

$$\Xi(\omega) = \int_{-\infty}^{\infty} \xi(t) e^{-j\omega t} dt$$

es imposible de realizar, principalmente, porque la integral (1.18) implica cumplir las condiciones de Dirichlet (sección §1.2.1), las cuales no ocurren para cualquier observación de proceso estacionario en el sentido amplio, particularmente, la condición de convergencia exige que el proceso sea absolutamente integrable; condición que sólo se alcanzaría para señales aleatorias de contenido cero.

El empleo de la TF exige la modificación de la representación para las observaciones de la señal estacionaria, de tal manera que la integral (1.18) converja, para lo cual, la forma

más sencilla consiste en el truncamiento de la trayectoria  $\xi(t)$ , empleado en el cálculo de sus momentos de una observación sobre un intervalo de tiempo  $(-T, T)$ , descrito en (3.11a) y (3.11b), que usa la función ventana rectangular:

$$\xi_T(t) = \text{rect}_T(t) \xi(t) \quad (3.21)$$

La trayectoria de la señal aleatoria truncada de (3.21) es válida, mientras la varianza del proceso sea finita. Así, se asegura la convergencia de la integral (1.18).

De otra parte, el empleo del teorema de Parsevall, mostrado en la Tabla 1.2, para la observación truncada (3.21), asumiendo su valor medio igual a cero, da como resultado:

$$\int_{-T}^T \xi_T^2(t) dt = \frac{1}{2\pi} \int_{-\infty}^{\infty} |\Xi_T(\omega)|^2 d\omega$$

cuyo promedio se obtiene acorde con la ventana de estimación (3.10):

$$\frac{1}{2T} \int_{-T}^T \xi_T^2(t) dt = \frac{1}{4\pi T} \int_{-\infty}^{\infty} |\Xi_T(\omega)|^2 d\omega$$

La parte izquierda de la última expresión es proporcional a la potencia del proceso en el intervalo de análisis  $(-T, T)$ . Es más, cuando se analizan las señales aleatorias estacionarias, al hacer el intervalo  $T \rightarrow \infty$ , la potencia tiende al valor cuadrático medio dado en (3.11a), esto es,

$$\begin{aligned} E \left\{ \frac{1}{2T} \int_{-T}^T \xi_T^2(t) dt \right\} &= E \left\{ \frac{1}{4\pi T} \int_{-\infty}^{\infty} |\Xi_T(\omega)|^2 d\omega \right\} \\ \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T E \{ \xi_T^2(t) \} dt &= \lim_{T \rightarrow \infty} \frac{1}{4\pi T} \int_{-\infty}^{\infty} E \{ |\Xi_T(\omega)|^2 \} d\omega \\ \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T \overline{\xi_T^2(t)} dt &= \lim_{T \rightarrow \infty} \frac{1}{4\pi T} \int_{-\infty}^{\infty} E \{ |\Xi_T(\omega)|^2 \} d\omega \\ \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T \overline{\xi_T^2(t)} dt &= \frac{1}{2\pi} \int_{-\infty}^{\infty} \lim_{T \rightarrow \infty} \frac{E \{ |\Xi_T(\omega)|^2 \}}{2T} d\omega \\ E \{ \overline{\xi_T^2(t)} \} &= \frac{1}{2\pi} \int_{-\infty}^{\infty} \lim_{T \rightarrow \infty} \frac{E \{ |\Xi_T(\omega)|^2 \}}{2T} d\omega \end{aligned}$$

En el caso de los procesos ergódicos, el promedio de ensamble es igual al promedio de

tiempo, con lo que se obtiene,

$$\overline{\xi_T^2(t)} = \frac{1}{2\pi} \int_{-\infty}^{\infty} \lim_{T \rightarrow \infty} \frac{E \left\{ |\mathcal{E}_T(\omega)|^2 \right\}}{2T} d\omega \quad (3.22)$$

El operador dentro de la integral (3.22) es el promedio de tiempo del espectro de la observación de la señal aleatoria, denominado *densidad espectral de potencia* (DEP),

$$S_\xi(\omega) \triangleq \lim_{T \rightarrow \infty} \frac{E \left\{ |\mathcal{E}_T(\omega)|^2 \right\}}{2T} \quad (3.23)$$

Si la señal  $\xi(t)$  tiene unidades  $[V]$ , entonces la densidad  $S_\xi(\omega)$  tiene unidades  $[V^2/Hz]$ , que en correspondencia con (3.22), determina el valor cuadrático medio del proceso. Así,

$$\overline{\xi_T^2(t)} = \frac{1}{2\pi} \int_{-\infty}^{\infty} S_\xi(\omega) d\omega \quad (3.24)$$

La densidad espectral de potencia puede ser interpretada como la potencia media concentrada en los límites de una banda de frecuencia con banda de paso igual a  $1 Hz$  para una frecuencia central del espectro igual a  $\omega/2\pi [Hz]$ .

Atendiendo a las condiciones de realización físicas de los procesos aleatorios, la DEP tiene, entre otras, las siguientes propiedades:

- (a).  $S_\xi(\omega) \in \mathbb{R}$ , está determinada en el espacio de los reales,
- (b).  $0 \leq S_\xi(\omega) < \infty$ , es positiva semidefinida y acotada,
- (c).  $S_\xi(\omega) = S_\xi(-\omega)$ , es par. En consecuencia, la representación de  $S_\xi(\omega)$  en forma de funciones racionales, debe contener estrictamente polinomios de potencias pares:

$$S_\xi(\omega) = \frac{s_0(\omega^{2n} + a_{2n-2} + \dots + a_2\omega^2 + a_0)}{\omega^{2m} + b_{2m-2} + \dots + b_2\omega^2 + b_0}, \quad S_\xi(\omega) \in \mathbb{Q}$$

La expresión (3.24) se obtiene asumiendo el valor medio igual a 0 del proceso aleatorio. Sin embargo, se puede demostrar que la relación puede ser generalizada, obteniéndose:

$$\sigma_\xi^2 = \frac{1}{2\pi} \int_{-\infty}^{\infty} S_\xi(\omega) d\omega \quad (3.25)$$

**Transformada de Wiener-Jinchin.** La expresión (3.11c) define la función de correlación como la esperanza matemática del producto de dos funciones determinadas en el tiempo. Sin embargo, de la sección anterior, resulta que la DEP está relacionada con la esperanza matemática del producto de las TF de estas misma funciones. Entonces, debe existir una relación directa entre esas dos esperanzas matemáticas.



Sea una señal aleatoria con DEP, definida en (3.23),

$$S_{\xi}(\omega) = \lim_{T \rightarrow \infty} \frac{E \{ |\mathcal{E}_T(\omega)|^2 \}}{2T} = \lim_{T \rightarrow \infty} \frac{E \{ \mathcal{E}_T(\omega) \mathcal{E}_T(-\omega) \}}{2T}$$

Desarrollando las respectivas integrales de Fourier, se obtiene

$$\begin{aligned} S_{\xi}(\omega) &= \lim_{T \rightarrow \infty} \frac{1}{2T} E \left\{ \int_{-T}^T \xi_T(t_1) e^{j\omega t_1} dt_1 \int_{-T}^T \xi_T(t_2) e^{-j\omega t_2} dt_2 \right\} \\ &= \lim_{T \rightarrow \infty} \frac{1}{2T} E \left\{ \int_{-T}^T dt_2 \int_{-T}^T e^{-j\omega(t_2-t_1)} \xi_T(t_1) \xi_T(t_2) dt_1 \right\} \end{aligned}$$

El operador de promedio, atendiendo a la propiedad de linealidad de la TF, se puede incluir dentro de la integral:

$$S_{\xi}(\omega) = \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T dt_2 \int_{-T}^T e^{-j\omega(t_2-t_1)} E \{ \xi_T(t_1) \xi_T(t_2) \} dt_1$$

El operador de promedio de las funciones dentro de la integral corresponde a la función de correlación propia  $R_{\xi}(t_1, t_2)$  para la señal estacionaria en análisis, pero truncada  $\xi_T(t)$ . Realizando el cambio de notación,  $\tau = t_2 - t_1$ , por lo que  $dt_2 = d\tau$ , la anterior expresión toma la forma:

$$\begin{aligned} S_{\xi}(\omega) &= \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T-t_1}^{T-t_1} d\tau \int_{-T}^T e^{-j\omega\tau} R_{\xi}(t_1, t_1 + \tau) dt_1 \\ &= \int_{-\infty}^{\infty} \left( \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T R_{\xi}(t_1, t_1 + \tau) dt_1 \right) e^{-j\omega\tau} d\tau \end{aligned} \quad (3.26)$$

La expresión dentro del operador de promedio en la integral (3.26) corresponde a la función de correlación (3.11c), asumiendo la media del proceso  $\overline{\xi(t)} = 0$ . En otras palabras, la DEP de la señal aleatoria  $\xi(t)$  corresponde a la TF de la respectiva estimación en el tiempo de la función de correlación propia  $R_{\xi}(t, t + \tau)$ ,

$$S_{\xi}(\omega) = \mathcal{F} \{ \overline{R_{\xi}(t, t + \tau)} \}$$

Por cuanto en los procesos estacionarios se cumple que la función de correlación no depende del tiempo inicial del intervalo de análisis, esto es,  $\overline{R_{\xi}(t, t + \tau)} = R_{\xi}(\tau)$ , entonces, se obtiene que la DEP de un proceso estacionario en el sentido amplio es la TF de su

función de correlación:

$$S_{\xi}(\omega) = \mathcal{F} \{R_{\xi}(\tau)\} \quad (3.27)$$

De manera inversa se puede obtener que:

$$\mathcal{F}^{-1} \{S_{\xi}(\omega)\} = \mathcal{F}^{-1} \{\mathcal{F} \{R_{\xi}(\tau)\}\} = R_{\xi}(\tau) \quad (3.28)$$

El par conjugado de expresiones (3.27) y (3.28) conforma la *Transformada de Winner-Jinchin* que tiene un valor fundamental en el análisis de señales estacionarias, debido a que establecen la relación de su representación entre el dominio del tiempo (la función de correlación) y el dominio de la frecuencia (la DEP).

**Ejemplo 3.9.** Hallar la DEP para la función de correlación  $R_{\xi}(\tau) = \sigma_{\xi}^2 \exp(-\alpha|\tau|)$ .

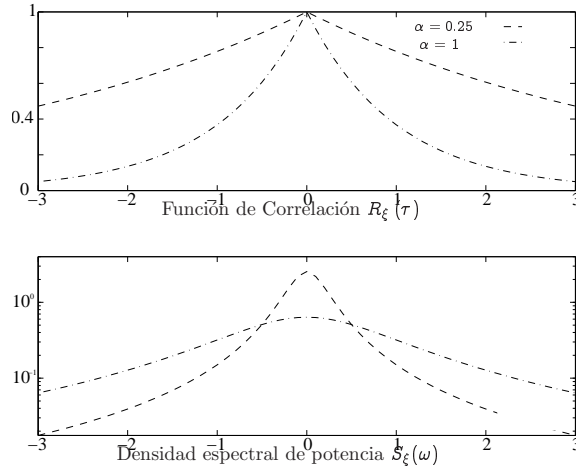
Empleando la relación (3.29a) se tiene:

$$S_{\xi}(\omega) = \frac{2}{\pi} \int_0^{\infty} \sigma_{\xi}^2 e^{-\alpha|\tau|} \cos \omega \tau d\tau = \frac{2\sigma_{\xi}^2}{\pi} \int_0^{\infty} e^{-\alpha\tau} d\tau \cos \omega \tau d\tau$$

La anterior integral se puede resolver por tabla:

$$S_{\xi}(\omega) = \frac{2\sigma_{\xi}^2}{\pi} \left( \frac{e^{-\alpha\tau}}{\alpha^2 + \omega^2} (\alpha \cos \omega \tau + \omega \sin \omega \tau) \right) \Big|_0^{\infty} = \frac{2\sigma_{\xi}^2}{\pi} \frac{\alpha}{\alpha^2 + \omega^2}$$

En la Figura 3.4 se presentan la función de correlación (parte superior) y la respectiva DEP (parte inferior); ambas calculadas para los casos de  $\alpha = 1$  y  $\alpha = 0.25$  y asumiendo un valor de la varianza unitario.



**Figura 3.4.** Resultados del ejemplo 3.9.

Empleando la propiedad de paridad de la función de correlación, las expresiones (3.27) y (3.28) se pueden llevar a las respectivas formas:

$$S_{\xi}(\omega) = 2 \int_0^{\infty} R_{\xi}(\tau) \cos \omega \tau d\tau \quad (3.29a)$$

$$R_{\xi}(\tau) = \frac{1}{\pi} \int_0^{\infty} S_{\xi}(\omega) \cos \omega \tau d\omega \quad (3.29b)$$

En la práctica, en calidad de valores de aleatoriedad para los procesos aleatorios, también se emplean los diferentes parámetros de su DEP, por ejemplo los siguientes:

– *Ancho de banda efectivo de espectro*

$$\Delta\omega_e = \frac{1}{S_{\xi}(0)} \int_{-\infty}^{\infty} S_{\xi}(\omega) d\omega$$

el cual está relacionado con el intervalo de correlación (3.15), por la expresión

$$\Delta\tau = \frac{1}{2} \frac{S_{\xi}(0)}{R_{\xi}(0)}$$

– Los siguientes momentos de frecuencia:

$$m_{1\omega} = \frac{2}{\sigma_{\xi}^2} \int_0^{\infty} \omega S_{\xi}(\omega) d\omega, \quad m_{2\omega} = \frac{2}{\sigma_{\xi}^2} \int_0^{\infty} \omega^2 S_{\xi}(\omega) d\omega, \quad \sigma_{\xi}^2 = \frac{2}{\sigma_{\xi}^2} \int_0^{\infty} (\omega - m_{1\omega})^2 S_{\xi}(\omega) d\omega$$

De otra parte, en el núcleo de la integral (3.26) está presente la función de correlación propia, que en el caso de considerar la función de correlación mutua (definida en (3.3b)), entonces, la función espectral respectiva es definida como la *densidad espectral de potencia mutua*:

$$S_{\xi\eta}(\omega) = \int_{-\infty}^{\infty} R_{\xi\eta}(\tau) e^{-j\omega\tau} d\tau \quad (3.30)$$

Cabe anotar, que cuando las señales aleatorias  $\xi(t)$  y  $\eta(t)$  tienen correlación nula, esto es,  $R_{\xi\eta}(\tau) = 0$ , la correspondiente DEP mutua también es de contenido 0.

La relación entre las densidades de potencia  $S_{\xi\eta}(\omega)$  y  $S_{\eta\xi}(\omega)$  se deduce a partir del análisis de las partes real e imaginaria de la definición (3.30),

$$S_{\xi\eta}(\omega) = \int_{-\infty}^{\infty} R_{\xi\eta}(\tau) \cos \omega \tau d\tau - j \int_{-\infty}^{\infty} R_{\xi\eta}(\tau) \sin \omega \tau d\tau, \quad (3.31)$$

donde las respectivas funciones de correlación mutua se definen como

$$R_{\xi\eta}(\tau) = \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T \xi(t) \eta(t + \tau) dt, \quad (3.32a)$$

$$R_{\eta\xi}(\tau) = \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T \eta(t) \xi(t + \tau) dt \quad (3.32b)$$

En la última integral, (3.32b), se realizan los siguientes cambios de variables:  $\lambda = t + \tau$ , luego,  $t = \lambda - \tau$ ,  $dt = d\lambda$ , con lo cual,

$$R_{\eta\xi}(\tau) = \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T \eta(\lambda - \tau) \xi(\lambda) d\lambda$$

regresando de nuevo a la variable  $t$ , se obtiene,

$$R_{\eta\xi}(\tau) = \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T \eta(t) \xi(t - \tau) dt$$

Al comparar la última expresión con la definición (3.32b) se observa el cumplimiento de la igualdad  $R_{\xi\eta}(\tau) = R_{\eta\xi}(-\tau)$ , que al sustituir en la (3.31), resulta en

$$S_{\eta\xi}(\omega) = \int_{-\infty}^{\infty} R_{\xi\eta}(-\tau) \cos \omega\tau d\tau - j \int_{-\infty}^{\infty} R_{\xi\eta}(-\tau) \sin \omega\tau d\tau$$

Sea  $-\tau = \lambda$ , entonces

$$\begin{aligned} S_{\eta\xi}(\omega) &= \int_{-\infty}^{\infty} R_{\xi\eta}(\lambda) \cos \omega\lambda d(-\lambda) - j \int_{-\infty}^{\infty} R_{\xi\eta}(\lambda) \sin(-\omega\lambda) d(-\lambda) \\ &= \int_{-\infty}^{\infty} R_{\xi\eta}(\lambda) \cos \omega\lambda d\lambda + j \int_{-\infty}^{\infty} R_{\xi\eta}(\lambda) \sin \omega\lambda d\lambda \end{aligned}$$

de lo cual, finalmente, se obtiene que

$$S_{\eta\xi}(\omega) = S_{\xi\eta}^*(\omega)$$

**Ruido blanco Gaussiano (RBG).** Se denomina *ruido blanco Gaussiano* al modelo de señal aleatoria que se define convencionalmente como una señal ergódica con FDP Gaussiana y DEP constante en todo el dominio de la frecuencia (Figura 3.5(b)),

$$S(\omega) = N_0, \omega \in (-\infty, \infty). \quad (3.33)$$

El modelo (3.33), denominado *ruido blanco* por su analogía espectral con la luz blanca, tiene función de correlación propia obtenida mediante la TF (Figura 3.5(a)):

$$R(\tau) = \int_{-\infty}^{\infty} \frac{N_0}{2} e^{j2\pi f\tau} df = \frac{N_0}{2} \delta(\tau) \quad (3.34)$$

Del modelo (3.34), se observa que  $R(\tau) = 0, \forall \tau \neq 0$ , de tal manera que cualquier par de diferentes valores tomados del RBG no son correlacionadas y, por lo tanto, estadísticamente son independientes.

Cabe anotar que el modelo (3.34) implica que la varianza del RBG  $\sigma^2 \rightarrow \infty$ .

En realidad, todos los dispositivos de procesos de señales eléctricas poseen un ancho de banda finito  $\Delta\omega$ , y en consecuencia cualquier clase de ruido también tendrá un ancho de banda finito a la salida de todo dispositivo. Cuando se asume la constancia de la DEP en el ancho de banda finito  $\Delta\omega < \infty$ , se tiene el siguiente modelo de ruido:

$$\begin{cases} S(\omega) = N_0, & (-\Delta\omega < \omega < \Delta\omega), \\ R(\tau) = N_0 \Delta\omega \operatorname{sinc}(2\Delta\omega\tau), \end{cases} \quad (3.35)$$

Puesto que el ancho de banda espectral resultante (3.35) es menor que del ruido blanco, al modelo filtrado se le conoce como *ruido rosado* o *blanco de banda finita*, para el cual además se cumple que su potencia de salida es finita e igual a  $N_0 \Delta\omega$  y su valores instantáneos están correlacionados.

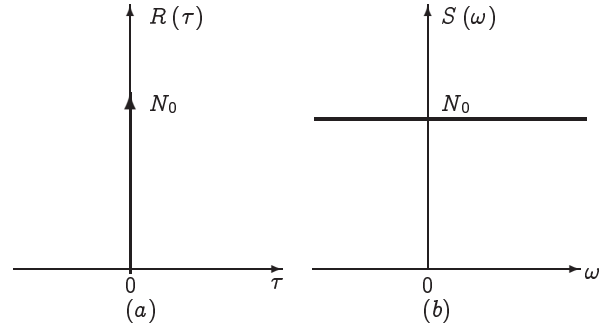
En la práctica, se define el *ancho de banda equivalente del ruido*  $\Delta\omega_e$ :

$$\Delta\omega_e = \frac{1}{H_0} \int_0^{\infty} |H(f)|^2 df, \quad (3.36)$$

donde  $H_0 = |H(f)|_{\max}$  es la ganancia de voltaje del filtro en la frecuencia central. Luego, la potencia de ruido promedio filtrado es:

$$N = \int_{-\infty}^{\infty} |H(f)|^2 N_0 df = N_0 \int_0^{\infty} |H(f)|^2 df = N_0 H_0 \Delta\omega_e,$$

La última ecuación muestra el efecto del filtro separado en dos partes: la selectividad de frecuencia relativa, representada por medio de  $\Delta\omega_e$ , y la ganancia de potencia representada por medio de  $H_0$ . Por definición, el ancho de banda equivalente de ruido de un filtro ideal es su ancho de banda real. En los filtros reales,  $\Delta\omega_e$  es algo mayor que el ancho de banda a 3 dB [14].



**Figura 3.5.** Modelo del ruido blanco Gaussiano

### 3.1.5. Convergencia y continuidad de procesos aleatorios

El análisis de modelos dinámicos, excitados por procesos aleatorios, se basa bien en la solución de ecuaciones diferenciales que describen los sistemas, bien en el análisis de otras características equivalentes. En cualquier caso, como ocurre en el estudio de excitaciones determinísticas, es necesario determinar el alcance de las definiciones de los conceptos de la continuidad, la diferenciabilidad e integrabilidad de los procesos aleatorios.

**Convergencia.** Sea la sucesión de valores aleatorios  $\{\xi_k : k = 1, \dots, n\}$ , para la cual, la convergencia se puede determinar en alguno de los siguientes sentidos [15]:

1. La sucesión  $\{\xi_k\}$  converge al valor aleatorio *con probabilidad de 1* (casi seguramente) si se cumple que,

$$P\{\xi_n \rightarrow \xi\}, \text{ para } n \rightarrow \infty$$

2. La sucesión  $\{\xi_k\}$  converge a  $\xi$  *en el sentido probabilístico*, si para cualquier  $\varepsilon > 0$ ,

$$P\{\xi_k - \xi > \varepsilon\} = 0, \text{ para } n \rightarrow \infty$$

3. La sucesión de valores aleatorios  $\{\xi_k\}$  converge a  $\xi$  en el *sentido cuadrático medio*, cuando,

$$\lim_{n \rightarrow \infty} E\{|\xi_n - \xi|^2\} = 0$$

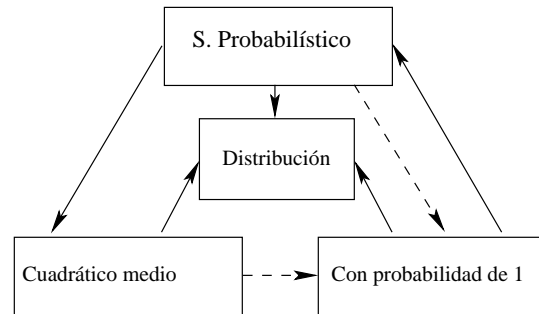
que frecuentemente se nota como,  $\text{l.i.m.}_{n \rightarrow \infty} \xi_n = \xi$ .

4. Se dice que  $\xi_n$  converge a  $\xi$  *por distribución*, si dadas las respectivas funciones de distribución,  $F_n(x)$  y  $F(x)$ ,  $\forall x$ , se cumple que  $F_n(x) \rightarrow F(x)$ , asumiendo  $n \rightarrow \infty$ .

Se considera que la convergencia con probabilidad de 1 implica la convergencia en el sentido probabilístico, mientras, la convergencia por valor cuadrático medio también implica la convergencia en el sentido probabilístico, como se muestra en la Figura 3.6, en la cual con líneas a trozos se representa la relación restringida a ciertas clases de sucesiones aleatorias.

En la práctica, el valor límite  $\xi$  no se conoce, entonces, similar al caso de sucesiones determinísticas como criterio de convergencia se emplea el *criterio de Cauchy*,

$$\lim_{n \rightarrow \infty} |\xi_{n+m} - \xi_n| \rightarrow 0, \forall m$$



**Figura 3.6.** Relación entre las definiciones de convergencia probabilística

Cuando el anterior límite existe, entonces se da la convergencia para cualquiera de sus formas anteriormente definidas. Las definiciones dadas de convergencia para sucesiones aleatorias se pueden extender a los procesos continuos aleatorios  $\{\xi(t), t \in T\}$ , teniendo en cuenta que las definiciones no se restringen solamente a una observación en particular del proceso, sino al conjunto de trayectorias en su totalidad.

**Continuidad.** Mediante las definiciones anteriores de convergencia, también es posible determinar la continuidad, así por ejemplo, una señal aleatoria  $\xi(t)$  se define como continua en  $t$  con probabilidad 1, cuando se cumple que:

$$\lim_{\tau \rightarrow 0} \xi(t + \tau) = \xi(t)$$

Sin embargo, no debe entenderse que la continuidad definida para los procesos aleatorios asegura la continuidad de cualquiera de sus trayectorias. Por ejemplo, un proceso de Poisson con FDP, dada por la expresión (3) del ejemplo 2.8, para un valor dado de  $t = t_0$ , es un proceso continuo, pero cualquiera de sus observaciones no lo es.

Con frecuencia, el criterio más aceptado de convergencia para un proceso aleatorio es el del valor cuadrático medio, el cual se asume que es un *proceso de segundo orden*, cuando se cumple que  $E\{\xi^2(t)\} < \infty, \forall t \in T$ .

Un proceso aleatorio de segundo orden,  $\xi(t), t \in T$ , es continuo en  $t$  en el sentido cuadrático medio, si se cumple que,

$$\lim_{\tau \rightarrow 0} E\{(\xi(t + \tau) - \xi(t))^2\} = 0 \quad (3.37)$$

Además, de (3.37) se obtiene que

$$\begin{aligned} E\{\xi(t + t_0), \xi(t)\} &= E\{(\xi(t + t_0) - m_{1\xi}(t) - \xi(t) - m_{1\xi}(t))^2 + \\ &\quad + 2(\xi(t + t_0) - \xi(t))((m_{1\xi}(t + t_0) - m_{1\xi}(t)) - (m_{1\xi}(t + t_0) - m_{1\xi}(t)))^2\} \\ &= R_\xi(t + t_0, t + t_0) - 2R(t + t_0, t + t_0) - 2R(t + t_0, t) + R(t, t) + \\ &\quad + (m_{1\xi}(t + t_0) - m_{1\xi}(t))^2 \end{aligned}$$

Por lo tanto, para la continuidad del proceso  $\xi(t)$  en  $t$ , es necesario que tanto su función de correlación  $R_\xi(t, t)$  como su valor medio  $m_\xi(t)$  sean funciones continuas. De otra parte, teniendo en cuenta que el valor cuadrático medio es positivo definido, entonces, la condición suficiente de continuidad corresponde a,

$$\begin{aligned} &E\{(\xi(t + t_0) - m_{1\xi}(t + t_0) - \xi(t) - m_{1\xi}(t))^2\} \\ &= R(t + t_0, t + t_0) - 2R(t + t_0, t) + R(t, t) \geq 0 \end{aligned}$$

**Diferenciabilidad.** Un proceso aleatorio  $\xi(t), t \in T$ , de segundo orden se define como diferenciable en el sentido cuadrático medio en el punto  $t_0 \in T$ , si se cumple que para el

límite

$$\lim_{\tau \rightarrow 0} \frac{\xi(t_0 + \tau) - \xi(t_0)}{\tau} = \xi'(t_0)$$

existe un valor  $\lim_{\tau \rightarrow 0} E \left\{ \left( \frac{\xi(t_0 + \tau) - \xi(t_0)}{\tau} - \xi'(t_0) \right)^2 \right\} = 0$ . Cuando esta condición se cumple para todo  $t \in T$ , entonces se dice que el proceso aleatorio es *diferenciable*.

La condición necesaria de diferenciabilidad del proceso en el punto  $t_0 \in T$  se halla analizando la siguiente sucesión de valores contiguos:

$$\begin{aligned} & E \left\{ \left( \frac{\xi(t_0 + \tau_1) - \xi(t_0)}{\tau_1} - \frac{\xi(t_0 + \tau_2) - \xi(t_0)}{\tau_2} \right)^2 \right\} \\ &= E \left\{ \frac{\xi(t_0 + \tau_1) - \xi(t_0)}{\tau_1} \frac{\xi(t_0 + \tau_1) - \xi(t_0)}{\tau_1} - 2 \frac{\xi(t_0 + \tau_1) - \xi(t_0)}{\tau_1} \frac{\xi(t_0 + \tau_2) - \xi(t_0)}{\tau_2} + \right. \\ &\quad \left. + \frac{\xi(t_0 + \tau_2) - \xi(t_0)}{\tau_2} \frac{\xi(t_0 + \tau_2) - \xi(t_0)}{\tau_2} \right\} = \\ &= \frac{R_\xi(t_0 + \tau_1, t_0 + \tau_2) - R(t_0 + \tau_1, t_0) - R(t_0, t_0 + \tau_2) + R_\xi(t_0, t_0)}{\tau_1 \tau_2} + \\ &\quad + \frac{m(t_0 + \tau_1) - m(t_0)}{\tau_1} \frac{m(t_0 + \tau_2) - m(t_0)}{\tau_2} \end{aligned} \quad (3.38)$$

Debido a la suposición sobre la diferenciabilidad del valor medio, la cual implica la existencia de la segunda derivada de la función  $R_\xi(t_1, t_2)$ , en los puntos de tiempo  $t_0$  y  $t_1 = t_2 - t_0$ , esto es,

$$\lim_{\tau_1, \tau_2 \rightarrow 0} E \left\{ \frac{\xi(t_0 + \tau_1) - \xi(t_0)}{\tau_1} \frac{\xi(t_0 + \tau_2) - \xi(t_0)}{\tau_2} \right\} = \frac{\partial^2 R(t_1, t_2)}{\partial t_1 \partial t_2} \Big|_{t_1=t_2=t_0} + m'(t_0)m'(t_0)$$

entonces, la parte derecha de la igualdad (3.38) es finita, para  $\tau_1, \tau_2 \rightarrow 0$ , se cumple que

$$E \left\{ \left( \frac{\xi(t_0 + \tau_1) - \xi(t_0)}{\tau_1} - \frac{\xi(t_0 + \tau_1) - \xi(t_0)}{\tau_1} \right)^2 \right\} \rightarrow 0$$

La condición suficiente de diferenciabilidad del proceso se halla asumiendo  $\tau_1 = \tau_2$ , con lo cual, la ec. (3.38) es la suma de dos términos positivos. Si la parte derecha de la igualdad tiende a 0, es suficiente que cada uno de sus términos de la derecha tiendan también a 0.

**Integrabilidad.** Sean un proceso aleatorio  $\xi(t)$  y la función determinística  $f(t)$ , tales que existan en el intervalo  $[a, b]$ , sobre el que se define la malla de puntos en el tiempo de la forma:  $a = t_0 < t_1 < \dots < t_n = b$ , sobre la cual se analiza la suma:

$$\mathcal{I}(n) = \sum_{i=1}^n f(t_i) \xi(t_i) (t_i - t_{i-1})$$



Si para un  $\max_{1 \leq i \leq n} (t_i - t_{i-1})$ , la suma converge en algún vector límite, que a propósito es otra variable aleatoria, entonces ese límite se denomina *integral del proceso aleatorio*  $\xi(t)$

$$\mathcal{J} = \int_a^b f(t)\xi(t)dt$$

La convergencia de las sumas  $\mathcal{J}(n)$  se analiza en el sentido cuadrático medio, esto es,  $E\{(\mathcal{J}(n) - \mathcal{J})^2\} \rightarrow 0$ ,  $n \rightarrow \infty$ , mientras el límite  $\mathcal{J}$  es la integral cuadrática media. Un proceso aleatorio de segundo orden,  $\xi(t)$ ,  $t \in [a, b]$  con  $m_{1\xi}(t)$  y  $R_\xi(t_1, t_2)$  es integrable por Riemann, cuando existen las integrales,

$$\int_a^b f(t)m_{1\xi}(t)dt, \int_a^b \int_a^b f(t_1)f(t_2)R_\xi(t_1, t_2)dt_1dt_2$$

en este caso, se tiene que

$$E \left\{ \int_a^b f(t)\xi(t)dt \right\} = \int_a^b f(t)m_{1\xi}(t)dt$$

$$E \left\{ \int_a^b \int_a^b f(t_1)f(t_2)\xi(t_1)\xi(t_2)dt_1dt_2 \right\} = \int_a^b \int_a^b f(t_1)f(t_2)R(t_1, t_2)dt_1dt_2 + \left( \int_a^b f(t)m_{1\xi}(t)dt \right)^2$$

Algunas veces, es necesario analizar la integral estocástica de Stieltjes,  $\int_a^b f(t)d\xi(t)$  que se define por el límite cuando  $\max_{1 \leq i \leq n} (t_i - t_{i-1}) \rightarrow 0$ , por la siguiente suma:

$$\sum_{i=1}^n f(t_i) (\xi(t_i) - \xi(t_{i-1}))$$

Es de anotar que el concepto de continuidad y diferenciabilidad de un proceso aleatorio no son equivalentes. La exigencia sobre la diferenciabilidad de un proceso aleatorio es más fuerte y restrictiva, que la continuidad del mismo. Por ejemplo, el proceso estacionario con función de correlación  $R_\xi(\tau) = \sigma_\xi^2(-\alpha|\tau|)$  es continuo, pero en general se demuestra que no es diferenciable.

### Problemas

**Problema 3.1.** Hallar el primer momento inicial  $m_{1\xi}(t)$  y la varianza  $\sigma_\xi^2(t)$  del proceso con FDP,

$$p(\xi, t) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{\xi^2}{2} e^{2\alpha t} + \alpha t\right)$$

**Problema 3.2.** Hallar la función de correlación y determinar la condición de estacionariedad de la señal aleatoria  $\eta(t) = y(t)\xi(t)$ , siendo  $y(t)$  una función no aleatoria.

**Problema 3.3.** Calcular los intervalos de correlación para las siguientes funciones de correlación  $R_\xi(\tau)$ : 1.  $a \exp(-\alpha|\tau|)$  2.  $a \exp(-\alpha^2\tau^2)$  3.  $a \exp(-\alpha^2\tau^2) \cos \omega_0\tau$  4.  $a \operatorname{sinc} \alpha\tau$

**Problema 3.4.** Hallar la función de correlación propia de la señal aleatoria periódica,

$$\xi(t) = \sum_{n=1}^{\infty} \frac{1}{n} \sin n\Omega t$$

**Problema 3.5.** Hallar la función de correlación  $R_\xi(\tau)$  de un proceso aleatorio estacionario con DEP dada así:

$$S_\xi(\omega) = \begin{cases} N/2, & |\omega_1| \leq \omega \leq |\omega_2| \\ 0, & \text{otros valores de } \omega \end{cases}$$

**Problema 3.6.** Hallar la DEP  $S_\xi(\omega)$  del proceso aleatorio  $\xi$ , cuya función de correlación es igual a

$$R_\xi(\tau) = \begin{cases} \sigma_\xi^2 \left(1 - \frac{\tau}{T}\right), & |\tau| \leq T \\ 0, & \text{otros valores de } t \end{cases}$$

**Problema 3.7.** Demostrar que para un proceso estacionario dado, el cambio de escala  $a$  en el argumento de la función de correlación, corresponde al siguiente cambio de escala en la DEP,

$$R_\xi(a\tau) \Leftrightarrow \frac{1}{a} S_\xi\left(\frac{\omega}{a}\right)$$

**Problema 3.8.** Demostrar que, en forma general, la DEP conjunta  $S_{\xi\eta}(\omega)$  no es una función par.

### 3.2. Análisis experimental de señales estacionarias

Sea una señal aleatoria y variable en el tiempo  $\xi(t)$ , entonces la mayoría de sus valores de aleatoriedad (momentos iniciales, centralizados, función de correlación, etc.), pueden ser también función del tiempo. Excepto, la densidad espectral de potencia que varía en términos de la frecuencia angular  $\omega = 2\pi f$ . El análisis experimental de la variable aleatoria  $\xi(t)$  para la estimación de alguna de sus valores de aleatoriedad  $\tilde{\theta}(t)$ , exige la realización de  $n$  experimentos y la obtención de una serie de observaciones  $\xi_1(t), \xi_2(t), \dots, \xi_n(t)$ . Como resultado se obtiene la estimación en la forma:

$$\tilde{\theta}(t) = g_{\theta}(\xi_1(t), \dots, \xi_n(t)) \quad (3.39)$$

**Tareas básicas de estimación.** En el análisis estadístico, de acuerdo a la información a priori dispuesta, se aceptan tres formas básicas de estructura de una señal  $\xi(t)$ ,  $t \in T$ , considerada estacionaria en el sentido amplio:

1.  $\xi(t) = \xi_0(t)$ ,  $t \in T$
2.  $\xi(t) = m_{\xi} + \xi_0(t)$ ,  $t \in T$
3.  $\xi(t) = \sum_{k=1}^r \theta_k a_k(t) + \xi_0(t)$ ,

donde  $\xi_0(t)$  es un proceso estacionario en el sentido amplio con media igual a cero.

Sea  $\xi_k(t)$ ,  $t \in T_a$  una trayectoria de la variable aleatoria  $\xi(t)$ , observada en el intervalo de análisis de tiempo  $T_a$ , donde  $T_a$  puede ser un segmento de tiempo  $T_a \in [a, b]$  para el caso de variables continuas, o una sucesión de valores definidos en un tiempo de observación  $T_a \in \{t_k, k = 1, \dots, n\}$ , para el caso de una variable aleatoria discreta.

En concordancia con lo anterior, se establecen las siguientes tareas básicas de estimación en los procesos estacionarios:

1. Estimación de la función de DEP,  $S_{\xi}(\omega)$ , para proceso  $\xi(t)$ ,
2. Conocida la DEP  $S_{\xi}(\omega)$  del proceso  $\xi_0(t)$ , se exige estimar su valor medio  $m_{\xi}$ ,
3. Conocida la DEP del proceso  $\xi_0(t)$ , se exige estimar los parámetros  $\theta_1, \dots, \theta_r$  de la regresión  $\sum_{k=1}^r \theta_k a_k(t)$ , cuando la función  $a_k(t)$  se asume conocida.

Sea  $\tilde{\theta} = g_{\theta}(\xi(t))$ ,  $t \in T_a$ , la estadística destinada a la solución de cualquiera de los tres casos básicos anteriores. De todas las posibles estadísticas  $\tilde{\theta}$  se escogen aquellas que presenten las propiedades expuestas en el numeral §2.2, en particular, las siguientes:

- (a). *Linealidad.* El funcional  $g_{\theta}(\cdot)$  debe ser lineal,
- (b). *Ausencia de sesgo.* Si la estimación del parámetro  $\theta$  se realiza mediante la estadística  $\tilde{\theta}$ , se exige que  $E\{\tilde{\theta}\} = \theta$ ,

- (c). *Consistencia.* La estadística  $\tilde{\theta}$  debe converger en el sentido probabilístico al valor  $\theta$ , en la medida en que se aumente el intervalo de observación,
- (d). *Efectividad.* La estadística  $\tilde{\theta}$  debe tener la menor de todas las varianzas posibles dentro de las estadísticas de una clase dada.

### 3.2.1. Estimación de momentos

**Estimación del valor medio.** En concordancia con la clase de estimadores lineales y sin sesgo, propuestos en (3.8) para los procesos ergódicos, se define el siguiente estimador:

$$\tilde{m}_{1\xi} = \overline{\xi(t)} = \int_a^b w_m(t) \xi(t) dt$$

donde la función  $w_m(t)$  es la función ventana, que cumple la condición (3.9).

La ausencia de sesgo del anterior estimador se comprueba de lo siguiente:

$$\begin{aligned} E \left\{ \overline{\xi(t)} \right\} &= E \left\{ \int_a^b w_m(t) \xi(t) dt \right\} = \int_a^b w_m(t) E \{ \xi(t) \} dt = \overline{\xi(t)} \int_a^b w_m(t) dt \\ &= \overline{\xi(t)} \end{aligned}$$

La precisión en la aproximación de la estimación,  $\tilde{m}_{1\xi} \approx \overline{\xi(t)}$ , se caracteriza por la respectiva varianza de la potencia de error descrita en (2.63) [23]:

$$\begin{aligned} E \left\{ \left( m_{1\xi} - E \left\{ \overline{\xi(t)} \right\} \right)^2 \right\} &= E \left\{ m_{1\xi}^2 \right\} - E \left\{ \overline{\xi^2(t)} \right\} \\ &= 2 \int_0^T R_\xi(\tau) \int_0^{T-\tau} w_m(t) w_m(t+\tau) dt d\tau \end{aligned} \quad (3.40)$$

La precisión de la estimación depende básicamente de la función ventana, cuya optimización por el criterio del mínimo error cuadrático medio conlleva al aumento significativo del costo computacional. Por esto, en la práctica se emplea la ventana más simple,

$$w_m(t) = \begin{cases} \frac{1}{T}, & 0 \leq t \leq T \\ 0, & 0 > t > T \end{cases} \quad (3.41)$$

con lo que la estimación de la media, para señales aleatorias continuas, es de la forma:

$$\overline{\xi(t)} = \int_0^T \xi(t) dt \quad (3.42)$$

La estimación del valor medio para una señal aleatoria discreta  $\{\xi[k] : k = 1, \dots, N\}$ , de

forma similar, se realiza discretizando la expresión (3.42),

$$\tilde{m}_{1\xi} = \sum_{k=1}^N \xi[k] w_m[k], \quad (3.43)$$

donde la función de peso discreta cumple la condición,  $\sum_{k=1}^N w_m[k] = 1$ . En los argumentos de la serie (3.43) se asume la normalización de las bases de tiempo,  $x[kT_d] = x[k]$ .

En la práctica, es frecuente el uso de la función ventana rectangular,  $w_m[k] = 1/N$ ,  $k = 1, \dots, N$ , para la cual, la varianza del error (3.40) toma la forma:

$$\sigma_\varepsilon^2 = \frac{\sigma_\xi^2}{N^2} \left( N + 2 \sum_{k=1}^{N-1} (N-k) \rho_\xi[k] \right)$$

donde  $R_\xi(\tau) = \sigma_\xi^2 \rho_\xi(\tau)$ . La relación  $\sigma_\varepsilon^2 / \sigma_\xi^2$  debe corresponder al mínimo valor posible, que es inversamente proporcional al tamaño  $N$  de la sucesión  $\xi[k]$ . Por lo tanto, el valor  $N$  corresponde al compromiso entre el costo computacional y la precisión del estimador.

**Estimación de la varianza.** Cuando el valor medio  $\overline{\xi(t)}$  del proceso aleatorio es 0, la estimación de la varianza coincide con el valor cuadrático medio y puede realizarse acorde con la definición (3.11a), teniendo en cuenta la clase de estimadores propuestos en (3.8),

$$\tilde{m}_{2\xi} = \overline{\xi^2(t)} = \int_0^T w_m(t) \xi^2(t) dt$$

Si el valor medio no es conocido, entonces la estimación de la varianza será

$$\tilde{\sigma}_\xi^2 = \int_0^T w_\sigma(t) \left( \xi^2(t) - \overline{\xi(t)} \right)^2 dt = \int_0^T w_\sigma(t) \left( \xi^2(t) - \int_0^T w_m(t) \xi(t) dt \right)^2 dt \quad (3.44)$$

Si se asume la ventana  $w_m$  de (3.41), además al exigir la constancia de la otra ventana,  $w_\sigma = \text{const}$ , para asegurar la ausencia de sesgo en (3.44), la función  $w_\sigma$  toma la forma

$$w_\sigma(t) = \frac{1}{T(1 - k_\sigma)} \quad (3.45)$$

El valor de la constante  $k_\sigma$  se determina experimentalmente [23].

**Estimación de la función de correlación.** Se tienen diferentes métodos de estimación de la función de correlación, uno de los más empleados es el siguiente:

$$\tilde{R}_\xi(\tau) = \int_0^{T-\tau} w_R(t, \tau) (\xi(t) - \tilde{m}_{1\xi}) (\xi(t + \tau) - \tilde{m}_{1\xi}) dt \quad (3.46)$$

donde  $w_R(t, \tau)$  es la ventana que se propone para la estimación de función de correlación.

Cuando se tiene que  $\bar{\xi}(t) = 0$ , la estimación (3.46) no presenta sesgo:

$$\begin{aligned} E \left\{ \tilde{R}_\xi(\tau) \right\} &= E \left\{ \int_0^{T-\tau} w_R(t, \tau) \xi_0(t) \xi_0(t + \tau) dt \right\} = \int_0^T w_R(t, \tau) E \left\{ \xi_0(t) \xi(t + \tau) \right\} dt \\ &= R_\xi(\tau) \end{aligned}$$

Además, la varianza de la estimación resulta ser igual a

$$\begin{aligned} \sigma_R^2 &= \int_0^{T-\tau} \int_0^{T-\tau} w_R(\lambda_1, \tau) w_R(\lambda_2, \tau) \times \\ &\quad \times \left( R_\xi^2(\lambda_1 - \lambda_2) + R_\xi(\lambda_1 - \lambda_2 - \tau) R_\xi(\lambda_1 - \lambda_2 + \tau) \right) d\lambda_1 d\lambda_2 \\ &= \int_{-(T-\tau)}^{T-\tau} w_R^2(\lambda, \tau) \left( R_\xi^2(\lambda) + R_\xi^2(\lambda - \tau) R_\xi^2(\lambda + \tau) \right) d\lambda \end{aligned}$$

De la anterior expresión se deduce que para determinar la varianza de la estimación para la función de correlación hay que conocer la misma señal aleatoria. Por esta razón, la determinación de la precisión de la estimación se puede realizar después de realizar el procesamiento, además, hay que tener en cuenta el hecho concreto de que la misma varianza es un proceso aleatorio. El desconocimiento a priori de la función de correlación, hace que en la mayoría de los casos no se plantee el problema de optimización de la función ventana de estimación  $w_R$ . Aunque, en el procedimiento de prueba de hipótesis, cuando se comprueba la pertenencia o no de una función de correlación a una clase dada, puede ser necesario la determinación de la ventana óptima de estimación.

En la práctica, la estimación de la función de correlación se realiza a partir de la ventana similar a la propuesta en (3.45):

$$\tilde{R}_\xi(\tau) = \frac{1}{(T-\tau)} \int_0^{T-\tau} (\xi(t) - m_{1\xi}) (\xi(t + \tau) - m_{1\xi}) dt \quad (3.47)$$

para la cual, se tiene el siguiente valor medio

$$\begin{aligned} E \left\{ \tilde{R}_\xi(\tau) \right\} &= R_\xi(\tau) - \frac{2}{T(T-\tau)} \int_0^T \left( 1 - \frac{\lambda}{T} \right) (\tau R_\xi(\tau) + T R_\xi(\lambda - \tau)) d\lambda + \\ &\quad + \frac{1}{T(T-\tau)} \int_0^T (T + \tau - 2\lambda) (R_\xi(\tau) + R_\xi(\lambda - \tau)) d\tau \end{aligned}$$

que muestra un sesgo en la estimación  $\tilde{R}_\xi$ , el cual se disminuye mediante el factor de corrección:  $1/(1 - f(R_\xi(\tau)))$ , donde la dependencia  $f(\cdot)$  se establece empíricamente.

De otra parte, la estimación del índice de correlación  $\rho(\tau)$ , (2.46), se realiza mediante la relación de los valores,  $\tilde{\rho}_\xi(\tau) = \tilde{R}_\xi(\tau) / \tilde{R}_\xi(0)$ , que resulta ser asintóticamente no sesgada:

$$E \{ \tilde{\rho}_\xi(\tau) - \rho_\xi(\tau) \} \approx \frac{2}{T} \int_{-\infty}^{\infty} \left( \rho^2(t) \rho(\tau) - \rho(t) \rho(t-\tau) \right) dt$$

En realidad, la estimación del índice de correlación de un proceso ergódico, inclusive siendo conocido el valor medio del proceso, presenta un sesgo, cuyo valor disminuye en la medida en que aumenta la longitud de la observación de la señal aleatoria en análisis.

Finalmente, en el caso de tener sucesiones estacionarias aleatorias, la estimación de la función de correlación se toma de la siguiente forma:

$$\tilde{R}_\xi[k] = \frac{1}{N-k} \sum_{l=1}^{N-k} \xi_0[l] \xi_0[l-k] \quad (3.48)$$

### 3.2.2. Estimación de los coeficientes en la descomposición K-L

En la sección §3.1.3 se plantea la búsqueda de un conjunto óptimo de funciones base con dimensión  $p$ , dado en  $L^2(T)$ , para la representación discreta finita de la forma (1.4) a partir de diversas observaciones de un proceso aleatorio  $\xi(t)$ ,  $t \in T$ , de tal manera que la norma del error en  $L^2(T)$ , promediada sobre el conjunto de observaciones, sea la menor posible.

En particular, de (3.18) se tiene que:

$$\begin{aligned} \bar{\varepsilon}_T^2 &= \int_T E \{ \xi(t) \xi^*(t) \} dt - \sum_{i=1}^p \int_T \int_T E \{ \xi(t) \xi^*(t) \} \phi_i^*(t) \phi_i(\tau) dt d\tau \\ &= \int_T R_\xi(t, \tau) dt - \sum_{i=1}^p \int_T \int_T R_\xi(t, \tau) \phi_i^*(t) \phi_i(\tau) dt d\tau \end{aligned} \quad (3.49)$$

El primer término de (3.49) no depende del conjunto base  $\phi = \{\phi_i\}$ , por lo que la tarea se resume a hallar las  $p$  funciones ortogonales que maximizan el segundo término:

$$\sum_{i=1}^p \int_T \int_T R_\xi(t, \tau) \phi_i^*(t) \phi_i(\tau) d\tau = \sum_{i=1}^p \langle \mathcal{A}_\xi \{ \phi_i(t) \}, \phi_i(t) \rangle$$

que corresponde a una suma de funcionales cuadráticos. La condición de valor máximo se encuentra, al tener en cuenta que el núcleo del operador integral  $\mathcal{A}_\xi$ , definido como,

$$\mathcal{A}_\xi \{ \phi(t) \} = \int_T R_\xi(t, \tau) \phi(\tau) d\tau \quad (3.50)$$

que corresponde a la función de correlación propia de un proceso con valor medio cuadrático finito. El operador (3.50) tiene las siguientes propiedades:

- (a). Integración cuadrática del núcleo,  $\int_T \int_T |\mathcal{A}(t, \tau)|^2 dt d\tau < \infty$ , en particular, en forma de un operador del tipo Hilbert-Schmidt.
- (b). Simetría,  $\mathcal{A}(t, \tau) = \mathcal{A}^*(t, \tau)$ , que resulta de las propiedades de la función de correlación propia dada en la sección §3.1.2,  $R_\xi(t, \tau) = R_\xi^*(\tau, t)$ .
- (c). Naturaleza positiva semidefinida,  $\langle \mathcal{A}\{\phi\}, \phi \rangle = E \{ \|\xi, \phi\|^2 \} \geq 0$ .

De las anteriores propiedades resulta lo siguiente:

1. Los valores propios de la solución (3.50) conforman una sucesión cuadrática sumatoria con valores reales positivos y los cuales se pueden disponer en su orden de decrecimiento:

$$\lambda_1 \geq \lambda_2 \geq \dots, \geq \lambda_i \geq \lambda_{i+1} \geq \dots$$

2. El núcleo se puede representar por una serie monótona convergente, que incluye las funciones propias  $\psi = \{\psi_i\}$  del operador  $\mathcal{A}_\xi$ , de la forma:

$$R_\xi(t, \tau) = \sum_{i=1}^{\infty} \lambda_i \psi_i(t) \psi_i^*(\tau) \quad (3.51)$$

3. Las funciones propias pueden ser ortonormales, de tal forma que se cumpla (1.5), esto es,

$$\langle \mathcal{A}_\xi \{\psi_i\}, \psi_k \rangle = \langle \lambda_i \psi_i, \psi_k \rangle = \lambda_i \delta_{ik} \quad (3.52)$$

En general, en [24] se demuestra que al emplear el operador  $\mathcal{A}_\xi$  para cualquier conjunto base  $\{\phi_i\}$ , es cierto que,

$$\sum_{i=1}^p \langle \mathcal{A}_\xi \{\phi_i\}, \phi_i \rangle \leq \sum_{i=1}^p \langle \mathcal{A}_\xi \{\psi_i\}, \psi_i \rangle = \sum_{i=1}^p \lambda_i$$

La expresión (3.49) para el conjunto base óptimo hallado toma la forma,

$$\bar{\varepsilon}_T^2 = \int_T R_\xi(t, \tau) dt - \sum_{i=1}^p \langle \mathcal{A}_\xi \{\psi_i\}, \psi_i \rangle \quad (3.53)$$

en la cual, al reemplazar (3.51) y (3.52) se obtiene que:

$$\bar{\varepsilon}_T^2 = \sum_{i=1}^{\infty} \lambda_i \langle \psi_i, \psi_i \rangle - \sum_{i=1}^p \lambda_i = \sum_{i=p+1}^{\infty} \lambda_i \quad (3.54)$$

De lo anteriormente expuesto, se puede concluir que el subespacio con dimensión  $p$ , definido en  $L^2(T)$  y que es óptimo para la representación de un proceso aleatorio en  $t \in T$ ,



está dado por las  $p$  funciones propias de la ecuación

$$\int_T R_\xi(t, \tau) \psi_i(\tau) d\tau = \lambda_i \psi_i(t) \quad (3.55)$$

las cuales corresponden a los  $p$  mayores valores de  $\lambda_i$ . Por cierto, debido a que el error cuadrático medio de representación  $\bar{\epsilon}_T^2$  corresponde a la suma residual de los valores propios, entonces el error de aproximación se puede variar aumentando o disminuyendo el índice inferior  $p$  de la respectiva suma en (3.54). Así, en general, aumentando  $T$ , los valores propios también aumentan, luego es necesario una mayor cantidad  $p$  de términos de la descomposición para obtener la precisión deseada.

La descomposición de señales aleatorias que tienen función de correlación continua por la serie (3.16),  $\xi(t) \approx \sum_{i=1}^n \xi_i \psi_i(t)$ ,  $t \in T$ , en la cual el conjunto base óptimo de representación (3.53) son las funciones propias, corresponde a la descomposición *no correlacionada* de Karhunen-Loève (ver ec. (3.16)), en la que los coeficientes (3.17)  $\{\xi_i\}$  son ortogonales:

$$\begin{aligned} E\{\xi_m \xi_l^*\} &= E\{\langle \xi(t), \psi_m(t) \rangle \langle \xi(t), \psi_l(t) \rangle^*\} = \int_T \int_T R_\xi(t, \tau) \psi_m^*(t) \psi_l(t) dt d\tau \\ &= \lambda_l \delta_{ml} \end{aligned}$$

De otra parte, si se tiene un proceso aleatorio con valor medio  $\overline{\xi(t)} = 0$ , entonces los coeficientes, también tendrán valor medio  $\{\overline{\xi_i} = 0, \forall i\}$  y tiene correlación nula (son linealmente independientes).

Aunque la descomposición K-L brinda la menor cantidad de elementos en la representación de procesos aleatorios por medio de la serie (1.4), para un valor dado de error  $\bar{\epsilon}_T^2$ , sin embargo su empleo está limitado principalmente por las siguientes razones: la función de correlación de cualquier proceso aleatorio, en la mayoría de los casos, es desconocida y el procedimiento de solución de (3.55) no es conocido en forma general.

Usualmente, en calidad de funciones base se emplean las funciones ortogonales del tipo Fourier, polinomios de Chebyshev, Lagrange, Funciones de Walsh y Haar, entre otras. No obstante, las funciones base exponenciales de Fourier (1.8),  $\psi_i(t) = e^{j\omega_0 t}$ , son óptimas para la representación de procesos aleatorios estacionarios cíclicos  $\xi(t)$ , para los cuales se cumple,

$$E\{\xi(t + kT)\} = E\{\xi(t)\} = m_{1\xi}(t), \quad R_\xi(t_1 + kT, t_2 + mT) = R_\xi(t_m, t_2)$$

En este caso, los coeficientes de descomposición  $\{\xi_i\}$  determinadas como

$$\xi_k(t) = \frac{1}{T} \int_0^T \xi(t) e^{-jk\omega_0 t} dt \quad (3.56)$$

tienen correlación 0 y su varianza en la expresión (3.20), teniendo en cuenta (3.53) y (3.55),

está dada por el valor:

$$\sigma_{\xi_i}^2 = \int_T R_{\xi}(\tau) \exp(-ji\omega_0\tau) d\tau$$

Las funciones base de Fourier no son óptimas para procesos aleatorios diferentes a los cíclicos estacionarios. Por lo tanto, los coeficientes respectivos de descomposición serán correlacionados. Sin embargo, para procesos ergódicos, cuando  $T \rightarrow \infty$  los coeficientes de Fourier resultan tener correlación nula [15] y la transformada de Fourier se aproxima a la descomposición K-L. Además, se puede demostrar que la relación entre el error de representación y el número mínimo  $p$  de elementos necesarios en la serie (3.16) al emplear la funciones base de Fourier, está dada por [22],

$$\bar{\epsilon}_{T(F)}^2 \approx \frac{2\sigma_{\xi}^2}{\pi^2 p}$$

De otra parte, el sistema de representación de los polinomios de Chebyshev son óptimos en el sentido del criterio de aproximación uniforme de la representación, cuando se cumple que el valor

$$\max \left| \xi^{(k)}(t) - \sum_{n=0}^{N-1} \xi_n T_n(t) \right|$$

es cercano al mínimo posible, que pueda ser obtenido por algún polinomio de aproximación uniforme polinomial [22], donde  $\xi^{(k)}$  es la  $k$  observación del proceso.

El cálculo de los coeficientes  $\xi_k$  para la representación de las señales aleatorias usando los Polinomios de Chebyshev es difícil, debido al valor que toma la función de peso definida en (1.9)  $w(t) = 1/\sqrt{1-t^2}$ . En este sentido es preferible el uso de polinomios con peso constante  $w(t) = 1$ , por ejemplo los polinomios de Legendre, que aunque brindan un valor de aproximación peor que los polinomios de Chebyshev, en cambio ofrecen mayor comodidad en el cálculo de los respectivos coeficientes de representación.

En otro caso, el sistema de funciones de Walsh  $\{\text{wal}_n(t) : n = 0, 1, \dots, 2^m - 1\}$  se considera óptimo en el sentido de representación de señales aleatorias diádico-estacionarias [22], cuya función de correlación cumple la condición  $R(t, \tau) = R(t \oplus \tau)$ . Esta clase de ruido puede ser obtenido si a la entrada del filtro de Walsh se introduce RBG. En otras palabras, la transformada de Walsh de una señal aleatoria diádico-estacionaria se puede identificar con la transformada K-L.

### 3.2.3. Estimación de la densidad espectral de potencia

La DEP  $S_{\xi}(\omega)$ , de una parte, corresponde a la varianza de la descomposición espectral del proceso estacionario, tal y como se muestra en (3.25). Pero de otra parte, la DEP corresponde a la transformada inversa de la función de correlación (3.27). Por lo tanto, la estimación de la DEP se puede realizar mediante ambas formas de representación. Al

analizar la DEP como la densidad de la varianza espectral, la estimación sobre una trayectoria de longitud finita del proceso se hace a través de la descomposición en la serie de Fourier (3.16) de la señal aleatoria para  $\omega = \omega_k$ , esto es,

$$\tilde{S}_\xi(\omega) = \frac{1}{2T} (\alpha_k^2 + \beta_k^2) = \frac{1}{T} \left| \int_0^T \xi(t) e^{-j\omega t} dt \right|^2 \quad (3.57)$$

donde  $\alpha_k = \int_0^T \xi(t) \cos 2\pi kt / T dt$  y  $\beta_k = \int_0^T \xi(t) \sin 2\pi kt / T dt$ .

Debido a la ortogonalidad de la representación y al tomar el valor límite  $T \rightarrow \infty$ , entonces, ambos valores  $\alpha_k$  y  $\beta_k$  no son correlacionados. Además, en los procesos con estructura Gaussiana,  $\alpha_k$  y  $\beta_k$  como se explicó en la sección §3.1.3, también tienen FDP Gaussiana y, por ende, la estimación de  $\tilde{S}_\xi(\omega)$ , haciendo  $T \rightarrow \infty$ , corresponde a la  $\chi^2$ -densidad, (2.31), con dos grados de libertad, con lo cual se demuestra que la media y la varianza de la estimación corresponden a:

$$E \{ S_\xi(\omega_k) \} = S_\xi(\omega_k) \quad (3.58a)$$

$$\sigma_S^2 = S_\xi^2(\omega_k) \quad (3.58b)$$

Mientras, el valor de la media muestra ausencia de sesgo, la varianza no tiende a cero, motivo por la cual, la estimación (3.57) no se considera consistente.

La segunda forma de representación, a partir de (3.27) (empleando la estimación (3.47)), implica la siguiente forma de cálculo para la DEP:

$$\tilde{S}_\xi(\omega) = \int_{-T}^T e^{-j\omega\tau} \tilde{R}_\xi(\tau) d\tau = \int_{-T}^T \left( 1 - \frac{|\tau|}{T} \right) R_\xi e^{-j\omega\tau} d\tau \quad (3.59)$$

Sin embargo, el valor medio de (3.59) muestra un sesgo en la respectiva estimación:

$$E \{ \tilde{S}_\xi(\omega) \} = E \left\{ \int_{-T}^T \left( 1 - \frac{|\tau|}{T} \right) \tilde{R}_\xi e^{-j\omega\tau} d\tau \right\} = \tilde{S}_\xi(\omega) - 2 \int_T^\infty \cos \omega\tau R_\xi(\tau) d\tau$$

Por cuanto [3],

$$\lim_{v \rightarrow \infty} \int_a^b x(\zeta) \cos(v\zeta) d\zeta = \lim_{v \rightarrow \infty} \int_a^b x(\zeta) \sin(v\zeta) d\zeta = 0$$

entonces, para valores pequeño de  $T$ , se tienen distorsiones significativas en el valor de la estimación propuesta.

Sea la observación  $\{ \xi(t), T \in [a, b] \}$  para un proceso estacionario. De ambos métodos de estimación de la DEP, (3.57) y (3.59), se establece la estadística común, denominada

*periodograma* y definida por la relación:

$$s(\lambda; a, b) = \frac{1}{2\pi(b-a)} \left| \int_a^b \xi(t) e^{j t \lambda} dt \right|^2$$

que en el caso discreto, cuando  $\{\xi_k \in T : t_k, k = 1, \dots, n\}$ , toma la forma:

$$s(\lambda; n) = \frac{1}{2\pi n} \left| \sum_{t \in T} \xi(t) e^{j t \lambda} \right|^2 \quad (3.60)$$

Sin embargo, la estadística en (3.60) presenta las mismas limitaciones de inconsistencia dadas en (3.58a) y (3.58b), razón por la cual el periodograma se convierte en otro proceso aleatorio que, para valores grandes de  $n$ , presenta fuertes fluctuaciones de trayectoria (realizaciones con alto grado de disimilaridad). La ausencia de consistencia en las estimaciones propuestas de DEP ha generado la necesidad de empleo de diferentes métodos orientados a aumentar la efectividad de estimación por medio de diversas funciones de peso, que impliquen su alisamiento. En particular, para una observación dada  $\{\xi(t_k) : t_k \in T, k = 1, \dots, n\}$  se emplea la estadística,

$$\tilde{S}_\xi(\lambda) = \int_{-\pi}^{\pi} W_S(\lambda - \nu) s(\nu; n) d\nu \quad (3.61)$$

La función de peso  $W_S(\lambda)$  en frecuencia, (3.61), es similar a la ventana propuesta para la estimación en el tiempo, (3.8), por lo que se denomina *ventana espectral*, que cumple los requerimientos:

- (a). *Localización*.  $W_S(\lambda)$  debe tener un valor máximo en  $\lambda = 0$ ,
- (b). *Ausencia de sesgo*.  $\int_{-\pi}^{\pi} W_S(\lambda) d\lambda = 1$ ,
- (c). *Consistencia*.  $\sigma_S^2 \rightarrow 0$ , cuando  $n \rightarrow \infty$ . En particular, se considera el valor asintótico,

$$\lim_{n \rightarrow \infty} \frac{1}{n} \int_{-\pi}^{\pi} W_S^2(\lambda) d\lambda = 0$$

En la práctica es común el empleo de funciones de peso que se puedan representar en la clase:

$$W_n(\lambda) = 2 \sum_{-n+1}^{n+1} k_n(l) e^{-j \lambda l}$$

donde  $k_n(l) = k(l/m_n)$ , siendo  $\{m_n\}$  una sucesión creciente acotada de valores enteros,

<i>Ventana</i>	<i>Modelo</i>
Transformada finita de Fourier (Estimación de Daniell)	$W_S(\lambda) = \begin{cases} m_n, &  \lambda  \leq \pi/m_n \\ 0, &  \lambda  > \pi/m_n \end{cases}$ $\tilde{S}_\xi(\lambda) = \frac{m_n}{2\pi} \sum_{t=-n+1}^{n-1} \left(1 - \frac{ t }{n}\right) \tilde{R}_\xi(t) \frac{1}{\pi t} \sin \frac{\pi t}{m_n} e^{jt\lambda}$ $k(\nu) = \text{sinc}(\pi\nu)$
Estimación truncada	$W_S(\lambda) = 2 \sin\left(\frac{2m_n+1}{2}\lambda\right) \left(\sin \frac{\lambda}{2}\right)^{-1}$ $\tilde{S}_\xi(\lambda) = \frac{1}{2\pi} \sum_{t=-m_n}^{m_n} \left(1 - \frac{ t }{n}\right) \tilde{R}_\xi(t) e^{jt\lambda}$ $k(\nu) = \begin{cases} 1, &  \nu  \leq 1 \\ 0, &  \nu  > 1 \end{cases}$
Estimación de Bartlett	$W_S(\lambda) = \frac{\sin^2 \frac{m_n \lambda}{2}}{m_n \sin^2 \frac{\lambda}{2}}$ $\tilde{S}_\xi(\lambda) = \frac{1}{2\pi} \sum_{t=-m_n}^{m_n} \left(1 - \frac{ t }{n}\right) \left(1 - \frac{ t }{m_n}\right) \tilde{R}_\xi(t) e^{jt\lambda}$ $k(\nu) = \begin{cases} 1 -  \nu , &  \nu  \leq 1 \\ 0, &  \nu  > 1 \end{cases}$
Estimación de Tukey-Hamming	(ver estimación truncada) $\tilde{S}_\xi(\lambda) = \frac{1}{2} \tilde{S}_{\xi F}(\lambda) + \frac{1}{n} \tilde{S}_{\xi F}\left(\lambda - \frac{\pi}{m_n}\right) + \frac{1}{n} \tilde{S}_{\xi F}\left(\lambda + \frac{\pi}{m_n}\right)$ $k(\nu) = \begin{cases} (1 + \cos \pi\nu)/2, &  \nu  \leq 1 \\ 0, &  \nu  > 1 \end{cases}$

**Tabla 3.1.** Ventanas de estimación de densidad espectral de potencia

tales que  $m_n/n \rightarrow 0$ , cuando  $n \rightarrow \infty$ , donde  $k(x)$  es una función par acotada, que cumple las restricciones:

(a).  $k(0) = 1$ ,  $|k(\nu)| < 1$ ,  $\forall \nu < 1$ ,

(b). Además,  $\int_{-\infty}^{\infty} k^2(\nu) d\nu < \infty$ .

La Tabla 3.1 muestra ejemplos de ventanas espectrales empleadas en la estimación de la DEP. En calidad de estimación de la función de correlación  $\tilde{R}_\xi(t)$  se puede tomar la expresión (3.48).

### 3.2.4. Estimación de parámetros de regresión

Sea una señal aleatoria observada en la forma:

$$\xi(t) = \sum_{k=1}^r \theta_k a_k(t) + \epsilon_0(t) = x(t, \theta_1, \dots, \theta_r) + \epsilon_0(t)$$

donde  $\xi_0(t)$  es un proceso estacionario con valor medio cero,  $\{a_k : k = 1, \dots, r\}$  funciones determinísticas conocidas, asumidas como de carácter lineal,  $\{\theta_k : k = 1, \dots, r\}$  el conjunto de los parámetros desconocidos a estimar, siendo  $r$  el modelo del sistema (valor conocido). El término  $x(\cdot)$  de acuerdo a la clase de tareas de que resuelva, recibe el nombre de *señal útil* (comunicaciones, proceso de señales, etc.) o *tendencia* (en aplicaciones médicas, biológicas, sociológicas, etc.). Mientras, el término  $\epsilon_0(t)$  recibe el nombre de *ruido*.

En general, existen los siguientes métodos de estimación del vector  $\theta = (\theta_0, \theta_1, \dots, \theta_{m-1})$ , que describe los parámetros del modelo de dependencia lineal (2.78).

**Método de mínimos cuadrados.** A partir de la trayectoria  $x(t)$  del proceso estacionario  $\xi(t)$ , la estimación de  $\{\theta_k : k = 1, \dots, r\}$  se realiza minimizando el funcional, derivado del criterio de minimización de la potencia de error (2.63):

$$\int_a^b \left| x(t) - \sum_{k=1}^r \theta_k a_k(t) \right|^2 dt \quad (3.62)$$

Si a su vez, se tiene que  $\{a_k(t) \in L^2(a, b) : k = 1, \dots, r\}$ , entonces, se demuestra que

$$\tilde{\theta}_k = \sum_{l=1}^r c_{kl}^{-1} \int_a^b \overline{a_l(t)} x(t) dt$$

siendo  $c_{kl}^{-1}$  el elemento  $kl$  de la matriz, que corresponde a la inversa de la matriz con elementos:

$$c_{kl} = \int_a^b \overline{a_k(t)} a_l(t) dt$$

Una ventaja del método de los mínimos cuadrados está en que no exige el conocimiento de las propiedades, ni espectrales ni de correlación, del proceso  $\xi(t)$ . Además, no tiene sesgo, y en caso de asumir que el respectivo espectro de potencia de  $S_\epsilon(\lambda)$ , es acotado y positivo definido, se demuestra la consistencia de la estimación de  $\{\theta_k : k = 1, \dots, r\}$ , con la siguiente condición de integrabilidad:

$$\int_{-\infty}^{\infty} |a(t)|^2 dt < \infty$$

**Estimación lineal de mínimo sesgo.** Cuando se conoce la densidad espectral de energía  $F_\epsilon = \mathcal{F}\{\epsilon_0(t)\}$  y, por ende, su función de correlación  $R_\epsilon(\tau)$ , se puede asumir que las funciones  $a_k(t)$ , que forman la regresión base  $x(t)$ , son tales que el proceso  $\xi(t)$  permite su representación espectral en la forma:

$$\xi(t) = \int_{-\infty}^{\infty} e^{jt\lambda} d\zeta(\lambda)$$

para la cual se cumple que

$$d\zeta(\lambda) = \sum_{k=1}^r \theta_k \overline{\alpha_k(\lambda)} dF_\epsilon(\lambda) + d\zeta_0(\lambda)$$

donde  $d\zeta_0(\lambda)$  corresponde al proceso espectral

$$\epsilon_0(t) = \int_{-\infty}^{\infty} e^{jt\lambda} d\zeta_0(\lambda)$$

mientras  $\{\alpha_k(\lambda) : k = 1, \dots, r\}$  son funciones integrables en el sentido cuadrático sobre la medida espectral  $F_\epsilon(\cdot)$ :

$$\int_{-\infty}^{\infty} |\alpha_k(\lambda)|^2 dF_\epsilon(\lambda) < \infty$$

entonces,  $\{\alpha_k \in L^2(F_\epsilon) : k = 1, \dots, r\}$  corresponden a la solución de la ecuación integral de Wiener-Hopf:

$$\int_{-\infty}^{\infty} e^{jt\lambda} \alpha_k(\lambda) dF_\epsilon(\lambda) = a_k(t), \forall k \quad (3.63)$$

Cuando existe la solución de (3.63), la tarea de estimación del conjunto de parámetros de la regresión  $\{\theta_k : k = 1, \dots, r\}$  consiste en la solución del sistema algebraico de ecuaciones lineales.

## Problemas

**Ejercicio en el CP 3.1.** Simular una sucesión aleatoria  $\{\xi[n] : n = 0, \dots, N-1\}$ , para diferentes valores de longitud de trayectoria  $N = 2^m$ ,  $m = 4, \dots, 12$ , con FDP  $\mathcal{N}(0, 1)$ . Verificar de forma cualitativa la normalidad, de acuerdo con la estimación de los valores para los coeficientes de asimetría y exceso.

**Ejercicio en el CP 3.2.** Sea el proceso aleatorio  $\eta(t)$  compuesto por la suma de  $N$  señales estacionarias  $\xi_n(t)$ , estadísticamente independientes, todas con FDP Gaussiana  $\mathcal{N}(0, \sigma_\xi)$ . Hallar la media, varianza y FDP del proceso  $\eta(t)$ . Comprobar experimentalmente, para un valor de significación de  $\alpha = 0.05$  los resultados obtenidos y establecer el comportamiento del error de estimación de los momentos del proceso de acuerdo a la longitud de las trayectorias de las señales  $\xi_n(t)$ .

**Ejercicio en el CP 3.3.** Sea el proceso aleatorio  $\xi(t)$ , que se asume con FDP Gaussiana  $\mathcal{N}(0, \sigma_\xi)$  y función de correlación

$$R_\xi(\tau) = \sigma_\xi^2 \exp(-\alpha |\tau|) \left( \cos \omega \tau + \frac{\alpha}{\omega} \sin \omega |\tau| \right)$$

Hallar la probabilidad  $P\{\dot{\xi} > c\}$  de que el valor de la derivada  $d\xi/dt$  no exceda el valor  $c$ . Simular dependencia entre el valor encontrado de probabilidad encontrado y diferentes valores del coeficiente  $\alpha$ .

**Ejercicio en el CP 3.4.** A la entrada del circuito  $RC$ , mostrado en la Figura 1.3 en régimen de integrador, se tiene RBG  $\xi(t)$  con parámetros  $\mathcal{N}(m_{1\xi}, \sigma_\xi)$ . Hallar el tiempo  $t_c$ , cuando el valor medio del proceso aleatorio  $\eta(t)$  a la salida del integrador alcanza el valor  $c = 0.95m_{1\eta}$ . Hallar la dependencia de  $t_c$  con respecto al valor medio  $m_{1\xi}$  para diferentes valores de la constante de integración  $\tau = RC$ .

**Ejercicio en el CP 3.5.** Realizar el cálculo numérico de la varianza  $\sigma_\eta^2(t)$  de la integral del proceso aleatorio  $\xi(t)$  con función de correlación  $R_\xi(\tau) = \sigma_\xi^2 \exp(-\alpha |\tau|) (1 - \alpha |\tau|)$ , para diferentes valores de la constante de integración  $\alpha$ .

**Ejercicio en el CP 3.6.** A la entrada del circuito  $RC$ , mostrado en el ejemplo (1.3) en régimen de integrador, se tiene la señal aleatoria  $\xi(t)$  con función de correlación y DEP, respectivamente dadas por la expresiones

$$R_\xi(\tau) = \sigma_\xi^2 \exp(-\alpha |\tau|), \quad S_\xi(\omega) = \sigma_\xi^2 \frac{1/\alpha}{1 + (\omega/\alpha)^2}$$

Hallar las correspondientes funciones de correlación  $R_\eta(\tau)$  y DEP  $S_\eta(\omega)$  del proceso  $\eta(t)$  a la salida del integrador, así como su valor de la varianza a la salida del circuito,  $\sigma_\eta^2$ . Realizar la simulación de las expresiones asintóticas de  $R_\eta(\tau)$  para  $\tau_0 \gg \tau$  y  $\tau_0 \ll \tau$ , siendo  $\tau = RC$ . Simular el comportamiento del sistema cuando a la entrada se presenta un tren pulsos cuadrados como el representado en la Figura 3.3.

**Ejercicio en el CP 3.7.** A la entrada de un sistema se tiene una observación de un proceso aleatorio  $\xi(t)$  con FDP diferente a la Gaussiana. El proceso aleatorio de salida es discretizado, de tal forma que puede ser representado en forma aproximada por la suma  $\eta[n] = \sum_k h[k] \xi[n-k]$ . Hallar experimentalmente, para que condiciones la FDP del proceso a la salida puede ser considerada normal.



### 3.3. Estimación espectral no paramétrica

#### 3.3.1. Método de los periodogramas

Dada una sucesión inicial de valores discretos, formada de la trayectoria de señal aleatoria  $\{x[k] : x \in \xi\}$  con longitud  $N$ , a partir de la expresión (3.60), el cálculo del periodograma, mediante técnicas de proceso digital se puede escribir como:

$$s_x[k] = \frac{1}{N} |X[k]|^2 = \frac{1}{N} \left| \sum_{n=0}^{N-1} x[n] e^{-jnk} \right|^2, \quad X[k] = \sum_{n=0}^{N-1} x[n] e^{-jnk} \quad (3.64)$$

Luego, el periodograma se puede calcular empleando la TRF, como se muestra en el procedimiento `Periodograma`. La ec. (3.64) es la forma no modificada de cálculo del periodograma. Sin embargo, como se observa de (3.58b), la operación (3.64) no resulta en una estimación consistente. Con el fin de obtener estimaciones consistentes de la DEP, sobre observaciones discretas de longitud finita, se emplea el suavizado de la estimación (3.64) en frecuencia o tiempo.

Así, la expresión (3.64) se escribe como

$$s_x[k] = \frac{1}{N} \left| \sum_{n=-\infty}^{\infty} w_s[n] x[n] e^{-jnk} \right|^2$$

donde  $w_s[n]$  es la ventana rectangular.

No obstante, como antes se dijo, es posible el empleo de funciones ventana diferentes a la rectangular. El periodograma, calculado mediante una función diferente a la ventana rectangular, se denomina *modificado*, el cual se describe como

$$s_x[w; k] = \frac{1}{NV} \left| \sum_{n=-\infty}^{\infty} w[n] x[n] e^{-jnk} \right|^2$$

donde  $V = \left| \sum_{n=0}^{N-1} w[n] \right|^2 / N$  es un factor de normalización que asegura que la estimación del histograma, asintóticamente, no tenga sesgo. El procedimiento `Periodograma modificado` realiza el periodograma para diferentes ventanas.

```
% Periodograma
function sx = periodogram(x,n1,n2)
    if nargin == 1,
        n1 = 1;
        n2 = length(x);
    end;
    sx = abs(fft(x(n1:n2),1024)).^ 2;
    sx = sx/(n2 - n1 + 1);
    sx(1) = sx(2);
end;
```

```
%Periodograma modificado
function sx = mper(x,win,n1,n2)
    if nargin == 2,
        n1 = 1;
        n2 = length(x);
    end;
    N = n2-n1+1; w = ones(N,1);
    if win==2,
        w = hamming(N);
    elseif win==3, w = hanning(N);
    elseif win==4, w = bartlett(N);
    elseif win==5, w = blackman(N);
    end;
    xw = x(n1:n2).*w/norm(w);
    sx = N*periodogram(xw);
```

Como se explicó para la expresión (3.61), la introducción de una función ventana genera el alisamiento en la frecuencia del histograma, con lo cual, la estimación de la DEP es,

$$E\{s_x[w; k]\} = \frac{1}{2\pi NV} |s[k] * W[k]|^2 \quad (3.65)$$

donde  $W[k]$  es la TF de la función ventana  $w[n]$ . De (3.65), se observa que el nivel de alisamiento en el periodograma depende de la forma de la función  $w[n]$  empleada en el preproceso de la sucesión de entrada. Así, por ejemplo la ventana rectangular tiene el lóbulo principal muy angosto, con relación a la mayoría de ventanas conocidas, lo cual genera menor grado de alisamiento espectral, pero en cambio, esta ventana presenta el mayor nivel de lóbulos laterales, que pueden llevar al enmascaramiento de componentes espectrales de poca amplitud. Cabe anotar, que la estimación mediante el histograma modificado presenta las mismas limitaciones, descritas en (3.58a) y (3.58b), sin embargo, el empleo de las ventanas permite establecer un cierto balance entre la resolución espectral (ancho de banda del lóbulo principal) y la pérdida espectral (área de los lóbulos laterales).

**Promedio de histogramas.** Como consecuencia de la ausencia de sesgo en la operación

$$\lim_{N \rightarrow \infty} E\{s_x[k]\} = S_x[k]$$

se asume que al hallar una estimación consistente para  $E\{s_x[k]\}$ , la misma se mantiene consistente para  $S_x[k]$ . Luego, partiendo del hecho de que la operación de promediado, para un conjunto de observaciones no correladas de la señal aleatoria, conlleva a la estimación consistente del promedio de la variable aleatoria  $E\{\xi\}$ , entonces, se analiza la estimación de la DEP mediante el promedio de los periodogramas o *método de Bartlett*.

Sea  $x_i[n] \in \xi$ ,  $n = 0, \dots, L-1$ ,  $i = 1, 2, \dots, K$ , un conjunto de observaciones no correlacionadas de la señal aleatoria  $\xi(t)$ . Si a la trayectoria  $i$  le corresponde el histograma  $s_x[i, k]$ , entonces el valor medio del conjunto de histogramas será

$$\tilde{S}_x[k] = \frac{1}{K} \sum_{i=1}^K s_x[i, k]$$

cuya esperanza matemática es igual a

$$E\{\tilde{S}_x[k]\} = E\{s_x[i, k]\} = \frac{1}{2\pi} S_x[k] * W_B[k]$$

donde  $W_B(\lambda)$  es la TF de la ventana de Bartlett  $w[n]$ ,  $-L < m < L$ . Por lo tanto, como en el caso del periodograma, el valor medio  $\tilde{S}_x[k]$ , asintóticamente no presenta sesgo. Además, asumiendo la correlación 0 de las trayectorias  $x_i[n]$ , la varianza de la estimación  $\tilde{S}_x[k]$  es:

$$\text{var}\{\tilde{S}_x[k]\} = \frac{1}{K} \sum_{i=1}^K \text{var}\{s_x[i, k]\} \approx \frac{1}{K} S_x^2[k]$$

que tiende a cero, cuando  $K \rightarrow \infty$ . Luego, la estimación  $\tilde{S}_x[k]$  es consistente.

En la práctica, disponer de una cantidad amplia de trayectorias de un mismo proceso es muy difícil, sin embargo, es más frecuente disponer de una trayectoria de longitud suficientemente larga  $N$ . Entonces se puede generar un conjunto de periodogramas obtenidos de la división de la trayectoria inicial en  $K$  segmentos sin traslapo, empleando técnicas de análisis en tiempo corto, cada uno de ellos de longitud  $L$ , de tal manera que  $K < L < N$ .

El procedimiento **Periodograma de Bartlett** ilustra el cálculo de los periodogramas de Bartlett. En general, la estimación promediada de histogramas tiene la forma

$$\tilde{S}_{xB}[k] = \frac{1}{N} \sum_{i=1}^K \left| \sum_{n=0}^{L-1} x[iL + n] e^{-jnk} \right|^2$$

Se puede considerar que la estabilidad de la estimación  $\tilde{S}_{xB}[k]$  mejora, al disminuir la cantidad de segmentos de análisis de tiempo corto. Dado el valor  $N = KL$  se observa la relación de compromiso entre una alta resolución (cuando se tiene el máximo valor de  $L$ ) y de mínima varianza de la estimación (para el máximo valor posible de  $K$ ).

#### Promedio de histograma modificado.

El análisis de tiempo corto mediante segmentos sin traslapo puede generar discontinuidades en los valores contiguos de estimación. Por esta razón, es preferible el empleo de segmentos de traslapo que aseguren la suavidad en la estimación. De otra parte, el empleo de funciones ventana puede atenuar el efecto de la pérdida espectral, así como disminuir el sesgo de la estimación, sin embargo, a costa de una pérdida en la resolución espectral. Así, sea el valor de traslapo  $D$ , para segmentos de análisis de tiempo corto con longitud  $L$ , luego la trayectoria  $i$  tiene la forma

$$x[i, n] = x[n + iD], \quad n = 0, 1, \dots, L - 1$$

Al tomar  $K$  segmentos que cubran la totalidad del registro de longitud  $N$ , entonces, se tiene que  $N = L + D(K - 1)$ , por lo que la estimación del periodograma de Welch se escribe

```
%Periodograma de Bartlett
function sx = bart(x,nsect)
L = floor(length(x)/nsect);
sx = 0; n1 = 1;
for i=1:nsect,
    pe = periodogram(x(n1:n1+L-1));
    sx = sx + pe/nsect;
    n1 = n1 + L;
end;
```

```
%Periodograma de Welch
function sx = welch(x,L,over,win)
if (over>=1) | (over<0),
    error('Valor traslapo inválido');
end;
n1 = 1; n0 = (1-over)*L;
nsect = 1 + floor((length(x)-L)/n0);
sx = 0;
for i=1:nsect,
    mp = mper(x,win,n1,n1+L-1)/nsect;
    sx = sx + mp;
    n1 = n1 + n0;
end;
```

en la forma:

$$\tilde{S}_{xW}[w; k] = \frac{1}{KLV} \sum_{i=0}^{K-1} \left| \sum_{n=0}^{L-1} w[n] x[n + Di] e^{-jnk} \right|^2$$

que en términos del histograma modificado es igual a:

$$\tilde{S}_{xW}[w; k] = \frac{1}{K} \sum_{i=0}^{K-1} s[w; i, k]$$

con lo cual, el valor esperado del periodograma de Welch resulta en:

$$E \{ \tilde{S}_{xW}[k] \} = E \{ s_x[w; k] \} = \frac{1}{2\pi LV} S_x[k] |W[k, L]|^2$$

donde  $W[k, L]$  es la TDF de la función ventana con longitud  $L$ , usada en la estimación del periodograma de Welch. Por cuanto, el factor de normalización es

$$V = \left| \sum_{n=0}^{L-1} w[n] \right|^2 / L$$

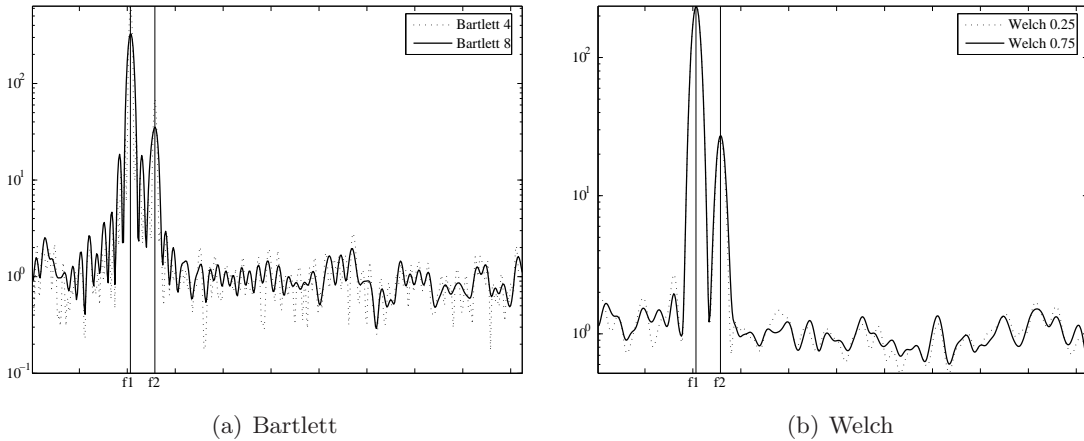
entonces, el periodograma de Welch también conlleva a una estimación sin sesgo.

---

**Ejemplo 3.10.** Sea la señal compuesta de dos armónicos y ruido blanco Gaussiano

$$x[n] = a_1 \sin(nk_1) + a_2 \sin(nk_2) + w[n]$$

dados los valores,  $k_1 = 0.2\pi$ ,  $k_2 = 0.25\pi$ ,  $a_1 = \sqrt{10}$ ,  $a_2 = 1$ . Hallar la estimación por Bartlett de la DEP.



**Figura 3.7.** Ejemplo de los periodogramas

En la Figura 3.7(a) se muestran los resultados obtenidos asumiendo los valores  $N = 512$  y

$L = 64$  para dos casos de análisis  $K = 4$  y  $K = 8$ . Como se observa, aunque la varianza de la estimación disminuye, al aumentar la longitud del segmento de tiempo corto  $K$ , empeora la resolución, condicionada por el aumento del lóbulo principal.

**Ejemplo 3.11.** En la Figura 3.7(b) se analiza el efecto del traslapo del ejemplo 3.10, dados los valores 50 y 75 %, en la varianza y resolución espectral del periodograma de Welch.

El cálculo de la varianza en la estimación de Welch es más complejo, en la medida, en que no se puede asegurar la ausencia de correlación entre las trayectorias con traslapo, caso en el cual debe aumentar la dispersión de los valores de la estimación. Por lo tanto, el empleo del traslapo aumenta la cantidad de segmentos de análisis corto, que en un principio debe disminuir la varianza del histograma. No obstante, al aumentar el valor del traslapo, aumenta el costo computacional en un valor proporcional a  $K$ . Además, el consiguiente aumento del traslapo aumenta el valor de correlación entre los segmentos que puede neutralizar el efecto de aumento en la cantidad  $K$ . En la práctica, el valor compromiso de traslapo se escoge entre un 50 y 75 % [25].

**Suavizado de periodogramas en frecuencia.** A partir de la expresión (3.59), se tiene que el periodograma se calcula empleando la TF de la función de correlación propia, en particular:

$$\tilde{S}_x[k] = \sum_{m=-(N-1)}^{N-1} r_x[m] \exp[-jmk] \quad (3.66)$$

donde

$$r_x[m] = \begin{cases} \frac{1}{N} \sum_{n=0}^{N-m-1} x[n+m]x^*[n], & 0 \leq m \leq N-1 \\ \frac{1}{N} \sum_{n=-1}^{N-|m|-1} x^*[n+|m|]x[n], & -(N-1) \leq m \leq -1 \end{cases} \quad (3.67)$$

se define como el *correlograma* de la sucesión aleatoria  $x[n]$  con longitud  $N$ .

De la expresión (3.66), se observa que se emplea un número diferente de operaciones dependiendo del intervalo de correlación. Así, en los extremos se tiene una sola operación. Cuando el intervalo de análisis es  $N-1$ , entonces  $r_x[N-1] = x[N-1]x^*[0]/N$ . Luego, la varianza de la estimación de la función de correlación tendrá valores muy altos en la medida en que crece hacia el valor extremo  $N$  del intervalo de correlación.

La estimación del correlograma, (3.67), implica el empleo de una ventana rectangular, no obstante, se puede variar el tipo de función ventana,  $w_r$ , tal que compense el efecto de las estimaciones no confiables en los extremos del intervalo de correlación, aunque esto implique la disminución del intervalo de proceso útil, y de esta forma, la disminución de la resolución. Además, se aumenta el fenómeno de pérdida espectral, debido a la aparición de lóbulos laterales, con lo cual la estimación tiende a presentar sesgo. Con el objeto de disminuir la pérdida espectral, se puede emplear el suavizado del periodograma, mediante

su convolución con una ventana espectral adecuada o *método de Blackman-Tukey*, que corresponde a la generalización del método de estimación por correlograma para la función de correlación propia, sopesado por una ventana. En este caso, la estimación de la DEP se puede describir como:

$$\tilde{S}_{xBT}[k] = \sum_{m=-M}^M w_r[m] r_x[m] e^{-jmk}$$

que en la frecuencia tiene la forma

$$\tilde{S}_{xBT}[k] = \frac{1}{2\pi} s_x[k] * W[k] \quad (3.68)$$

De (3.68), se ve que la estimación de Blackman-Tukey realiza el suavizado del periodograma mediante su convolución con la TDF de la ventana de correlación  $w_r[m]$ , la cual en la estimación de los correlogramas, debe cumplir las condiciones:

1.  $0 \leq w_r[m] \leq w_r[0] = 1$
2.  $w_r[-m] = w_r[m]$
3.  $w_r[m] = 0, \forall |m| > M, M \leq N - 1$

De la expresión (3.68), se observa que es suficiente, mas no necesaria, la condición de que el periodograma tenga un valor semidefinido positivo, que corresponde a la restricción  $W(\lambda) \geq 0$ , con  $-\pi \leq \lambda \leq \pi$ .

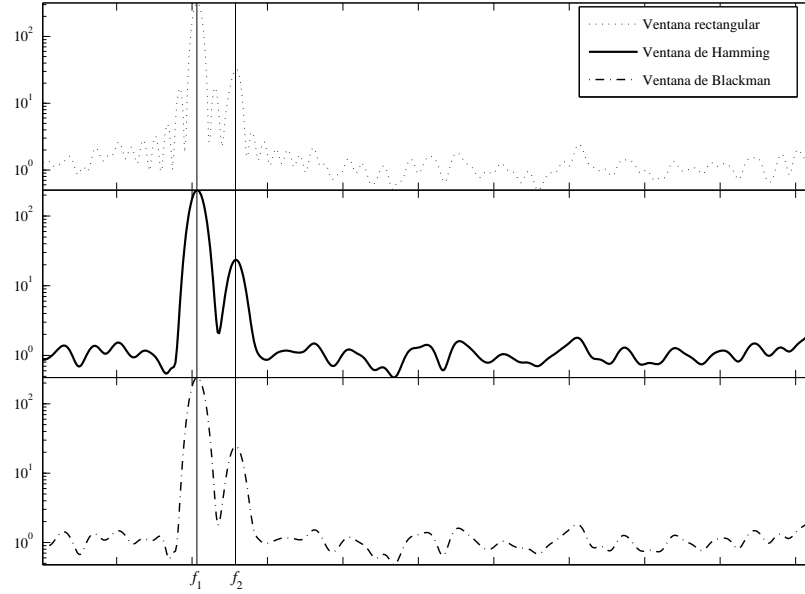
La ventana triangular de Bartlett cumple esta condición, mas no ocurre con las ventanas de Hamming y rectangulares, comunes en el proceso de señales. En la Figura 3.8 se muestra un ejemplo del cálculo del periodograma mediante el método de Blackman-Tukey, que presenta una varianza mucho menor, aunque empeore su resolución de frecuencia.

En general, se considera que la estimación de la DEP mediante cualquiera de los métodos de histograma, antes analizados, debe cumplir un compromiso entre la resolución espectral y la varianza de la estimación, por esto, la comparación de los mismos se realiza empleando las siguientes características estadísticas [25]:

$$(a). \nu = \frac{\text{var}\{\tilde{S}_x(\lambda)\}}{E^2\{\tilde{S}_x(\lambda)\}}$$

el cual corresponde al nivel de dispersión de la estimación y caracteriza la estabilidad estadística de la estimación.

```
% Periodograma de Blackman-Tukey
function sx = per_smooth(x,win,M,n1,n2)
if nargin==3, n1=1;
    n2 = length(x);
end;
R = cov(x(n1:n2),M);
r = fliplr(R(1,2:M),
R(1,1),R(1,2:M));
M = 2*M-1;
if win==2 w = hamming(N);
    elseif win==3
        w = hanning(N);
    elseif win==4 w = bartlett(N);
    elseif win==5 w = blackman(N);
end;
r = r'.*w;
sx = abs(fft(r,1024));
sx(1) = sx(2);
```



**Figura 3.8.** Ejemplo de periodograma de Blackman–Tukey

(b).  $\mu = \nu \Delta f$

donde  $\Delta f$  es la capacidad de resolución del método. El parámetro  $\mu$  permite determinar la resolución del DEP obtenida a partir de una trayectoria de longitud finita de la señal aleatoria, por cierto, a menor  $\mu$ , mayor calidad de la estimación.

Método	Estabilidad $\nu$	Resolución $\Delta f$	Índice $\mu = \nu \Delta f$
Periodograma	1	$0.89(2\pi/N)$	$0.89(2\pi/N)$
Bartlett	$1/K$	$0.89(2\pi/N)$	$0.89(2\pi/N)$
Welch	$(9/8)(1/K)$	$1.282(\pi/L)$	$0.72(2\pi/N)$
Blackman-Tukey	$(2/3)(M/N)$	$0.64(2\pi/M)$	$0.43(2\pi/N)$

**Tabla 3.2.** Comparación de los métodos de periodograma de estimación de DEP

En la Tabla 3.2 [25], se muestran las características estadísticas para los métodos del periodograma, descritos anteriormente. Se observa que cada uno de los métodos tiene más o menos el mismo valor compromiso de estabilidad y resolución, el cual es inversamente proporcional a la longitud  $N$  de la sucesión de valores discretos de la señal aleatoria. En general, sin importar que métodos se tengan, que ofrezcan mejor resolución o menor dispersión en la estimación, se cumple que el compromiso entre la resolución espectral y la estabilidad básicamente depende de la longitud,  $N$ , y de los valores discretos disponibles para el análisis de la sucesión aleatoria  $x[n]$ .

### 3.3.2. Algoritmo de cálculo del método de periodograma

Sean conocidos, para el proceso aleatorio en análisis, los valores del intervalo de discretización  $\Delta t$  y la resolución espectral necesaria  $\Delta f$  de análisis. Entonces, el intervalo de observación  $T$  y la respectiva cantidad  $N$  de valores discretizados de la señal se relacionan por la expresión [26]:

$$T = \frac{k_w}{\Delta f}, \quad N = \lfloor \frac{T}{\Delta t} \rfloor \quad (3.69)$$

donde  $k_w$  es un coeficiente que se determina de acuerdo al tipo de ventana empleada en la estimación.

El algoritmo del método de periodogramas se divide en dos etapas:

- el preproceso de la señal en el intervalo de observación,
- el promedio de los resultados durante varios intervalos de aproximación, con el fin de disminuir la dispersión de la estimación.

La primera etapa comprende los siguientes procedimientos:

1. Cálculo del valor  $N$ , a partir de (3.69). Por cuanto la TF se calcula mediante el algoritmo TRF, entonces, cuando  $N \neq 2^m$ ,  $m \in \mathbb{Z}$ , en cada una de las trayectorias, se realiza el relleno de ceros hasta el primer entero  $m$ , para el cual se cumpla la igualdad.
2. Selección de la función ventana  $w_s[n]$  y cálculo de la TRF, para cada una de las  $l$  trayectorias,  $x \in \xi$  disponibles  $\{x[n, k] : n = 0, \dots, N-1; k = l, \dots, L\}$ , que son alisadas por la respectiva ventana:

$$X[k, l] = \sum_{n=0}^{N-1} x[n, l] w_s[n] \exp(-j2\pi kn/N) \quad (3.70)$$

3. Cálculo del periodograma  $s[l, \lambda, N]$ , definido en (3.60), para cada una de las trayectorias  $l$ :

$$s[l, k, N] = \frac{|X[k, l]|^2}{\sum_{n=0}^{N-1} w_s[n]} \quad (3.71)$$

4. Dado un criterio de convergencia, si el valor  $s[l, k, N]$  no lo cumple, entonces, se deben repetir los pasos 1 y 2, sobremuestreando la sucesión de cada trayectoria en 2, 4, 8, ... veces, mientras no se disminuya adecuadamente la incertidumbre en la estimación del periodograma.

Una forma alterna de disminuir la incertidumbre está en realizar el procedimiento de rellenos de ceros aumentando también la longitud  $N$  de cada trayectoria en



$2, 4, 8, \dots, 2^m$  veces. En este caso, el arreglo de entrada de la TRF se modifica en la forma:

$$X[m; k, l] = \sum_{n=0}^{2^m N - 1} x[m; n, l] w_s[n] \exp\left(-j \frac{2\pi k n}{N 2^m}\right) \quad (3.72)$$

donde

$$x[m; n, l] = \begin{cases} x[n, l], & 0 \leq n \leq N - 1 \\ 0, & n > N - 1 \end{cases}$$

Al comparar las expresiones (3.70) y (3.72), se observa que  $X[m; k, l] = X[2^m k, l]$ , luego, la operación de compresión de escala de tiempo expande la envolvente espectral, pero no le cambia de forma. En la práctica, el relleno de ceros conlleva a la aparición de componentes adicionales entre los armónicos originales.

La segunda etapa comprende los siguientes procedimientos:

1. Selección del factor de traslapo  $\Delta$  entre los intervalos contiguos de observación. Usualmente, se toma un valor entre  $\Delta \in [0.5, 0.75]$ .
2. Cálculo del número total de intervalos de observación,  $N_i$ :

$$N_i = \lfloor ((N_T - \Delta N)) / (N - \Delta N) \rfloor$$

donde  $N_T$  es la longitud total del registro de la señal aleatoria en análisis.

3. Estimación de la DEP promediada:

$$\tilde{S}[k] = \frac{1}{N_i} \sum_{l=1}^{N_T} s[l, k, N] \quad (3.73)$$

4. Cálculo del coeficiente  $k_\sigma$ , que muestra la disminución en la varianza de la estimación de la DEP, debido a la operación de promediado, en cada uno de los intervalos de observación.

$$k_\sigma = \begin{cases} \left( \frac{1}{N_i} (1 + 2c^2(0.5)) - \frac{2}{N_i^2} c^2(0.5) \right)^{-1}, & \Delta = 0.5 \\ \left( \frac{1}{N_i} (1 + 2c^2(.75) + 2c^2(0.5)) - \frac{2}{N_i^2} (c^2(0.75) + 2c^2(0.5)) \right)^{-1}, & \Delta = 0.75 \end{cases}$$

donde los valores de  $c(\Delta)$  se determinan de acuerdo al tipo de ventana  $w_s[n]$  empleada en la estimación espectral [27].

### 3.3.3. Ventanas de estimación espectral

La multiplicación de los valores discretizados de la señal aleatoria  $x[n]$  por la función ventana,  $w_s[n]$  corresponde a la convolución en frecuencia de los respectivos espectros.

Entre las principales propiedades y características de las funciones ventana, empleadas en la estimación espectral, están las siguientes:

- Simetría,  $w[n] = w[N - n]$ ,  $n = 1, \dots, N - 1$ .
- Ancho de banda equivalente de ruido, que se determina como

$$\Delta f_R(N) = \frac{N \sum_{n=0}^{N-1} w^2[n]}{\left( \sum_{n=0}^{N-1} w[n] \right)^2}$$

a menor valor de  $\Delta f_R$ , menor es la potencia del ruido y, por lo tanto, menor distorsión que se puede causar sobre el proceso en análisis.

**Ejemplo 3.12.** Hallar el ancho de banda equivalente de ruido por la ventana triangular, cuando  $N = 2K$ , asumiendo que  $w[0] = 0$ .

La ventana triangular se determina como

$$w[n] = \begin{cases} n/K, & n = 1, \dots, K \\ (2K - n)/K, & n = K + 1, \dots, 2K - 1 \end{cases}$$

Entonces,  $\Delta f_R(N) = 2(2K^2 + 1)/(3K^2)$ , por cuanto en la práctica,  $N \gg 1$ , en lugar de  $\Delta f_R(N)$  se emplea su valor asintótico, cuando  $N \rightarrow \infty$ , esto es,  $\Delta f_R = \lim_{N \rightarrow \infty} \Delta f_R(N)$ . Particularmente, en el caso de la ventana triangular  $\Delta f_R = 1.33$ .

- Ancho de banda del lóbulo principal  $\Delta f_{LP}(N, g)$  de la magnitud del espectro, que típicamente se determina para un nivel fijo de atenuación  $g$ , dado en  $dB$ , con relación al valor máximo del módulo espectral,  $\Delta f_{LP}(N, g) = \gamma(N)N/\pi$ , donde  $\gamma(N)$  es la menor de todas las posibles raíces por módulo absoluto, obtenidas de la ecuación,  $|W(N)| = \max |W(N)|/10^{g/20}$ .

En la práctica, es usual el empleo del valor asintótico, dado el nivel  $g$ , para el ancho de banda del lóbulo principal

$$\Delta f_{LP}(g) = \lim_{N \rightarrow \infty} (\gamma(N)N/\pi)$$

Se considera que el valor  $\Delta f_{LP}(g)$  caracteriza la resolución del algoritmo de la TRF, para una función ventana dada, en particular, se asume que [28]

$$\kappa_o = \Delta f_{LP}(g)$$

- Ganancia coherente

$$G_c(N) = \frac{\sum_{n=0}^{N-1} w[n]}{\max |W(N)|}$$

con valor asintótico  $G_c = \lim_{N \rightarrow \infty} G_c(N)$ , que corresponde a la amplificación relativa del armónico de una señal, cuya frecuencia coincide con una de las frecuencias del conjunto base de la TRF.

- Máximo nivel de cualquiera de los lóbulos laterales,  $m_{LL}$ , obtenido para el método de TRF de la función ventana, medido en  $dB$  con relación al valor máximo del lóbulo principal. En este sentido, también se considera la velocidad de caída  $v_{LL}$  de los lóbulos laterales, medida en  $dB$  por octava (o década), que muestra qué tan rápido decrece la energía contenida en los lóbulos laterales. A mayor velocidad de caída, menor pérdida espectral presenta la ventana.
- Modulación de amplitud parásita,  $a_P$ , que caracteriza la amplitud relativa del armónico de la señal, después de su proceso mediante la función ventana y cálculo de la TRF, en el peor de los casos, cuando la frecuencia de la señal se encuentra exactamente en la mitad de un par de frecuencias base de la TRF.

El valor  $a_P$  se mide en  $dB$  y se define como

$$a_P = 20 \log |W(\omega)| / \max(W(\omega))$$

Los siguientes factores confluyen en la elección del tipo de función ventana:

- aplicación concreta del proceso de señales
- exigencias de costo computacional (recursos de proceso y tiempo de cómputo) en la solución de la aplicación

**Ejemplo 3.13.** Sea la sucesión

$$x[n] = \sum_k a_k \sin(2\pi nk/N + \phi_k), \quad n = 0, \dots, N-1 \quad (3.74)$$

donde  $a_k$  y  $\phi_k$  son los valores desconocidos para la amplitud y fase de los armónicos, respectivamente, los cuales coinciden en la frecuencia base de la TRF.

En este caso, el cálculo de  $a_k$  y  $\phi_k$  se puede realizar, simplemente empleando la función ventana rectangular. En particular, de (3.70) se tiene que

$$a_k = |X[k]|, \quad \phi_k = \arg(X[k])$$

Sin necesidad de recurrir al procedimiento de promedio, (3.73), en la medida en que no hay incertidumbre de medida.

Si en la expresión (3.74), asumiendo que se tiene un solo armónico, se agrega la perturbación  $\{r[n]\}$ , en forma de RBG, entonces para la estimación de los valores de  $a_k$  y  $\phi_k$  es necesario procesar la sucesión inicial mediante una función ventana, que provea el menor valor de  $\Delta f_R$ , por ejemplo, la ventana rectangular, y luego, se realiza el procedimiento de promediado (3.73).

**Ejemplo 3.14.** Sea la sucesión

$$x[n] = a_1 \sin(n\omega_1 + \phi_1) + a_2 \sin(n\omega_2 + \phi_2) \quad (3.75)$$

donde  $\omega_{11} \leq \omega_1 \leq \omega_{12}$ ,  $\omega_{21} \leq \omega_2 \leq \omega_{22}$ , siendo  $\omega_{ij}$ , las frecuencias base dadas inicialmente de la TRF. Hallar los valores de  $\omega_1$ ,  $\omega_2$ ,  $a_1$  y  $a_2$ , conocido que  $a_1 \approx a_2$ .

En este caso, el preproceso de la sucesión inicial se puede realizar mediante la función ventana rectangular. Asumiendo en (3.69)  $\Delta f = (\omega_{21} - \omega_{12})/2\pi$  y  $\kappa_o = \Delta_{LP}(g)$ , entonces se puede determinar el intervalo de observación  $T$ , con lo que conocido el intervalo de discretización  $\Delta t$  se puede hallar la longitud  $N_i$  de observación.

Sin embargo, el cálculo de la DEP, en correspondencia con (3.70) y (3.71) para  $k = 0, \dots, N-1$ , puede tener lugar una indeterminación que haga imposible estimar cada armónico debido a la presencia de varios valores iguales de  $\hat{S}(k)$ .

La forma directa para disminuir esta incertidumbre consisten en aumentar la longitud de la sucesión inicial agregando  $N(2^m - 1)$  ceros, con lo cual el espectro  $X[k]$  se calcula por la expresión (3.72). Inicialmente se prueba con  $m = 1$ . En caso de que la incertidumbre no permita aún el cálculo confiable de los armónicos, se aumenta sucesivamente  $m = 2, 3, \dots$  hasta que se obtenga un valor adecuado del armónico.

Debido a que la sucesión inicial se considera que no está perturbada, entonces, no hay necesidad de la operación de promedio (3.73).

## Problemas

**Problema 3.9.** Demostrar que la estimación de Bartlett para la DEP no presenta sesgo.

**Problema 3.10.** Demostrar que la función  $k(\nu)$  de la estimación de Tukey-Hamming para la DEP es cuadrado integrable,  $L^2(\mathbb{R})$ .

**Problema 3.11.** Demostrar que  $E\{\tilde{S}_x[k]\} = \frac{1}{2\pi} S_x[k] * W_B[k]$ .

**Problema 3.12.** Repetir el ejemplo 3.10 para la señal  $x[n] = 2 \sin(0.4\pi n) + 2 \cos(4.5\pi n + 0.5) + w[n]$ , con RBG con media 0 y varianza 0.2.

**Problema 3.13.** Demostrar que  $\tilde{S}_{xBT}[k] = \frac{1}{2\pi} s_x[k] * W[k]$ .

**Problema 3.14.** Considérese una sucesión aleatoria de longitud  $N$ , la cual ha de ser dividida en  $K$  segmentos, cada uno con  $M = N/K$  puntos. Si se conoce que la DEP tiene dos picos separados a una distancia de  $2[\text{rad/s}]$ , cual debe ser el mínimo valor de  $M$  para poder detectar correctamente la presencia de ambos picos?

**Problema 3.15.** Sea la sucesión aleatoria, cuya función de correlación propia se da por la expresión  $R[k] = 0.8^{|k|}$ ,  $k = 0, \pm 1, \pm 2, \dots$ . Si se emplean 10 valores discretos para estimar la función de correlación, mediante el correlogramas (3.67). Hallar el sesgo de la estimación para todo los  $k$ .

### 3.4. Estimación espectral paramétrica

En los métodos paramétricos, la señal aleatoria medida se analiza como la salida de un sistema lineal e invariante en el tiempo, al cual se aplica una entrada con características de aleatoriedad definidas a priori. En la práctica, el modelo más usado corresponde al excitado por RBG y con función de transferencia racional del sistema. Los procesos de salida de los modelos de esta clase tienen DEP que se describe completamente, mediante los coeficientes de la respectiva función de transferencia y la varianza del RBG de entrada.

En esencia, los métodos paramétricos incluyen alguna información sobre los procesos analizados, que tiene en cuenta sus propiedades y su naturaleza interna; información que se desprecia en los métodos no paramétricos. Como resultado, los métodos paramétricos permiten obtener estimaciones con mayor precisión y mayor resolución.

La estimación espectral paramétrica consta de tres etapas:

1. Selección del modelo paramétrico de la señal aleatoria analizada
2. Estimación de los parámetros del modelo seleccionado de la señal aleatoria de acuerdo a las trayectorias conocidas
3. Cálculo de la DEP mediante la sustitución en el modelo teórico de los parámetros estimados

#### 3.4.1. Reacción de un sistema lineal a una señal estacionaria

Sea la entrada  $x[n]$  una sucesión estacionaria con valor medio 0, entonces  $y[n]$  representa la salida del sistema lineal e invariante en el tiempo, la cual también es estacionaria y se describe como

$$y[n] = \mathcal{H}\{x[n]\}$$

La función de correlación propia de la respuesta se determina por la convolución [25]:

$$r_y[m] = r_x[m] * \left( \sum_{k=-\infty}^{\infty} h[k+m]h^*[k] \right)$$

que empleando la transformada  $\mathcal{Z}$  se obtiene

$$S_y(z) = S_x H(z) H^*(1/z^*) \quad (3.76)$$

donde  $S_x(z) = \mathcal{Z}\{r_x[m]\}$ ,  $S_y(z) = \mathcal{Z}\{r_y[m]\}$ ,  $H(z) = \mathcal{Z}\{h[m]\}$ .

Sea la entrada RBG, con valor medio igual a 0 y potencia  $\sigma_\eta^2$ , tal que

$$S_x(z) = \mathcal{Z}\{\sigma_\eta^2 \delta[m]\} = \sigma_\eta^2$$

entonces, la salida y la entrada del respectivo sistema lineal e invariante en el tiempo tiene

la forma:

$$S_y(z) = \sigma_\eta^2 H(z) H^*(1/z^*)$$

La última ecuación muestra la relación de proporcionalidad que hay entre la función de transferencia del filtro del modelo y la DEP de señal aleatoria analizada.

### 3.4.2. Modelos paramétricos de señales aleatorias

La descripción paramétrica de las estadísticas de segundo orden también aplica a las sucesiones de discretización de las señales aleatorias. En este sentido, el modelo es adecuado para la aproximación de señales aleatorias discretas, se describe mediante la señal de salida de un sistema lineal e invariante en el tiempo, expresada por la ecuación de diferencias en coeficientes complejos

$$x[n] = - \sum_{k=1}^p a_k x[n-k] + \sum_{k=0}^q b_k u[n-k] \quad (3.77a)$$

$$= \sum_{k=0}^{\infty} h[k] x[n-k] \quad (3.77b)$$

donde  $u[n]$  es la sucesión de excitación o la entrada del sistema,  $x[n]$  es la sucesión a la salida del filtro causal ( $h[k] = 0, \forall k < 0$ ), el cual conforma los datos observados. Se asume que el sistema lineal tiene función de transferencia racional,  $H(z) = B(z)/A(z)$ , para la cual los polinomios se determinan como

$$\begin{aligned} A(z) &= 1 + \sum_{k=1}^p a_k z^{-k} \\ B(z) &= 1 + \sum_{k=0}^q b_k z^{-k}, \quad b_0 = 1 \\ H(z) &= 1 + \sum_{k=1}^{\infty} h_k z^{-k} \end{aligned}$$

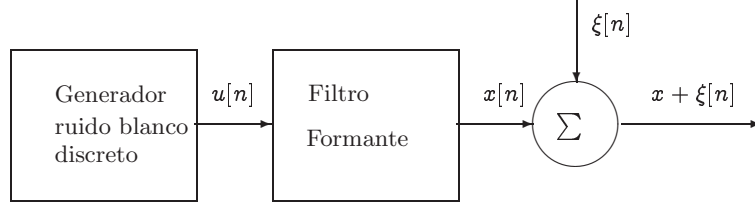
Se asume además que los ceros de los polinomios  $A(z)$  y  $B(z)$  se encuentran ubicados dentro del círculo unitario de convergencia del plano  $\mathcal{Z}$ , con el fin de garantizar la estabilidad del filtro.

En concordancia con (3.76) los espectros de entrada y salida en un sistema lineal e invariante en el tiempo se relacionan como:

$$S_x(z) = S_u(z) H(z) H^*(1/z^*) = S_u(z) \frac{B(z) B^*(1/z^*)}{A(z) A^*(1/z^*)} \quad (3.78)$$

La sucesión de entrada  $u[n]$ , usualmente, no es observable y, por lo tanto, no se puede emplear en el análisis espectral. Los modelos paramétricos, como se muestra en la Figura

3.9 emplean en calidad de señal de excitación una sucesión de RBG con valor medio cero y varianza  $\sigma^2$ , entonces,  $S_u(z) = \sigma^2$ .



**Figura 3.9.** Diagrama de formación de un modelo paramétrico de un proceso aleatorio

En general, las propiedades de la señal aleatoria modelada, dependen de la estructura y valores de los parámetros del filtro formante, además de las propiedades de la señal de entrada.

En caso de ser necesario el modelado de datos de medición en forma de una mezcla aditiva del modelo paramétrico más alguna perturbación  $\eta[n]$ , entonces, la perturbación se agrega a la salida del filtro formante, como se muestra en la Figura 3.9.

En la estimación de la DEP, se emplean modelos excitados por RBG los cuales se clasifican en tres clases

1. modelos de procesos autorregresivos (AR)
2. modelos de procesos de media deslizante (MA)
3. modelos de procesos autorregresivos y media deslizante (ARMA)

Las diferentes clases, que se muestran en la Tabla 3.3, se diferencian por el tipo de función de transferencia discreta del filtro formante y, en consecuencia, por el tipo de ecuación lineal de diferencia que describe la sucesión de salida aleatoria.

<i>Modelo paramétrico</i>	<i>Ecuación iterativa</i>	<i>Función de transferencia</i>
AR	$x[n] = - \sum_{k=1}^p x[n-k] + u[n]$	$\frac{1}{1 + \sum_{k=1}^p a_k z^{-k}}$
MA	$x[n] = u[n] + \sum_{k=1}^q b_k u[n-k]$	$1 + \sum_{k=1}^q b_k z^{-k}$
ARMA	$x[n] = \sum_{k=1}^p a_k x[n-k] + u[n] + \sum_{k=1}^q b_k u[n-k]$	$\frac{1 + \sum_{k=1}^q b_k z^{-k}}{1 + \sum_{k=1}^p a_k z^{-k}}$

**Tabla 3.3.** Características de los modelos paramétricos

**Modelos ARMA.** Se describen por la ecuación de diferencias generalizada (3.77a), en la cual como se observa de la Tabla 3.3, se tienen polos y ceros. La DEP para un modelo ARMA se obtiene al sustituir en (3.78),  $z = e^{j\Omega}$ :

$$S(\Omega) = \sigma^2 \left| \frac{B(\Omega)}{A(\Omega)} \right|^2 = \frac{\sigma^2 \mathbf{e}_q^H(\Omega) \mathbf{b} \mathbf{b}^H \mathbf{e}_q(\Omega)}{\mathbf{e}_p^H(\Omega) \mathbf{a} \mathbf{a}^H \mathbf{e}_p(\Omega)} \quad (3.79)$$

donde los vectores columna de los exponentes complejos  $\mathbf{e}_p$  y  $\mathbf{e}_q$ , mas los coeficientes  $\mathbf{a}$  y  $\mathbf{b}$  son de la forma

$$\begin{aligned} \mathbf{e}_p^\top &= (1, e^{j\Omega}, \dots, e^{jp\Omega}), & \mathbf{e}_q^\top &= (1, e^{j\Omega}, \dots, e^{jq\Omega}) \\ \mathbf{a}^\top &= (1, a, \dots, a_p), & \mathbf{b}^\top &= (1, b, \dots, b_q) \end{aligned}$$

Es anotar que la DEP se calcula en el rango de frecuencias normalizadas  $\Omega \in [-\pi, \pi]$ .

De la expresión (3.79), es claro que el modelo ARMA se caracteriza por los parámetros  $a$ ,  $b$  y la varianza del ruido blanco  $\sigma^2$ .

**Modelo MA.** De acuerdo con la función de transferencia mostrada en la Tabla 3.3, la respuesta a impulso en este caso es

$$x[k] = \begin{cases} b_k, & k = 0, 1, \dots, q; \\ 0, & \text{en otro caso.} \end{cases}$$

que tiene longitud finita y, por lo tanto, en concordancia con la expresión (1.57), los filtros formantes de los modelos MA pertenecen a la clase de los no recursivos. La DEP del proceso se calcula haciendo en (3.79)  $p = 0$ , de acuerdo a la expresión:

$$S(\Omega) = \sigma^2 |B(\Omega)|^2 = \sigma^2 \mathbf{e}_q^H(\Omega) \mathbf{b} \mathbf{b}^H \mathbf{e}_q(\Omega)$$

**Modelo AR.** De la Tabla 3.3, se observa que la respectiva función de transferencia no tiene ceros, ( $q = 0$ ), y presenta solamente polos (todo polos), los cuales tienen respuesta a impulso con longitud infinita, estos pertenecen a la clase de los recursivos. La DEP de un proceso AR, haciendo  $q = 0$ , resulta en

$$S(\Omega) = \frac{\sigma^2}{|A(\Omega)|^2} = \frac{\sigma^2}{\mathbf{e}_p^H(\Omega) \mathbf{a} \mathbf{a}^H \mathbf{e}_p(\Omega)}$$

### 3.4.3. Estimación de los parámetros a partir de la función de correlación

Sea conocida la función de correlación propia de la sucesión aleatoria en análisis. Entonces, al multiplicar ambas partes de la ecuación (3.77a) por el factor  $x^*[n - m]$  y calculando la



esperanza matemática se obtiene

$$E\{x[n]x^*[n-k]\} = -\sum_{l=1}^p a_l E\{x[n-l]x^*[n-k]\} + \sum_{l=0}^q b_l E\{u[n-l]x^*[n-k]\}$$

que en términos de la función de correlación se escribe como

$$r_x[k] = -\sum_{l=1}^p a_l r_x[k-l] + \sum_{l=0}^q b_l r_{ux}[k-l]$$

la función de correlación mutua  $r_{ux}[k]$  entre las secuencias de entrada y salida se pueden expresar mediante los parámetros de la respuesta impulso  $h[m]$ :

$$\begin{aligned} r_{ux}[k] &= E\{u[n+k]x^*[n]\} = E\{u[n+k]u^*[n] + \sum_{m=1}^{\infty} h^*[m]u^*[n-m]\} \\ &= r_u[k] + \sum_{m=1}^{\infty} h^*[m]r_u[k+m] \end{aligned}$$

por cuanto  $u[n]$  es una sucesión de RBG con varianza  $\sigma^2$ , entonces,

$$r_{ux}[k] = \begin{cases} 0, & k > 0; \\ \sigma^2, & k = 0; \\ \sigma^2 h^*[-k], & k < 0. \end{cases}$$

De lo cual se obtiene la expresión que relaciona los coeficientes del modelo ARMA con la función de correlación propia de la señal discretizada aleatoria  $x[n]$

$$r_x[k] = \begin{cases} r_x[-k], & k < 0; \\ -\sum_{l=1}^p a_l r_x[k-l] + \sigma^2 \sum_{l=k}^q b_l h^*[l-k], & 0 \leq k \leq q; \\ -\sum_{l=1}^p a_l r_x[k-l], & k > q. \end{cases} \quad (3.80)$$

donde  $h[0] = 1$ , por definición, además con el fin de cumplir la condición de causalidad, el filtro debe tener respuesta a impulso  $h[k] = 0, \forall k < 0$ .

Cabe anotar que la relación en los parámetros del modelo ARMA y la función de correlación es no lineal, en particular, el sistema de ecuaciones (3.80) es no lineal, debido a la presencia del término  $\sum_{l=k}^q b_l h^*[l-k]$ . Sin embargo, cuando el modelo paramétrico es del tipo AR, el modelo tiene carácter lineal. Así, asumiendo en (3.80)  $b_l = \delta[l]$ , entonces se obtiene

$$r_x[k] = -\sum_{l=1}^p a_l r_x[k-l] + \sigma^2 h^*[-k], \quad 0 \leq k \leq q$$

Debido a que  $h^*[-k] = 0, \forall k > 0$ ,  $h^*[0] = \lim_{z \rightarrow \infty} H(z) = 1$ , entonces de (3.80) se tiene que

$$r_x[k] = \begin{cases} -\sum_{l=1}^p a_l r_x[k-l], & k > 0; \\ -\sum_{l=1}^p a_l r_x[-l] + \sigma^2, & k = 0; \\ r_x^*[-k], & k < 0. \end{cases} \quad (3.81)$$

Las expresiones (3.81) corresponden a las ecuaciones normales de Yule-Walker para un proceso AR [29]. Estas ecuaciones, en general, caracterizan la relación no lineal entre los parámetros del proceso AR y la función de correlación propia de la señal aleatoria analizada. Sin embargo, conocida la función de correlación propia se pueden determinar los parámetros del modelo AR al resolver un sistema de ecuaciones lineales lo cual se puede observar al plantear (3.81) en forma matricial:

$$\begin{bmatrix} r_x[0] & r_x[-1] & \cdots & r_x[-(p-1)] \\ r_x[1] & r_x[0] & \cdots & r_x[-(p-2)] \\ \vdots & \vdots & \ddots & \vdots \\ r_x[p-1] & r_x[p-2] & \cdots & r_x[0] \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_p \end{bmatrix} = \begin{bmatrix} r_x[1] \\ r_x[2] \\ \vdots \\ r_x[p] \end{bmatrix} \quad (3.82)$$

De esta manera, conocida la sucesión de valores  $r_x[k]$ ,  $-p \leq k \leq p$ , para la función de correlación propia, entonces los parámetros del modelo AR se pueden hallar como la solución del sistema de ecuaciones lineales (3.82). Cabe anotar que la matriz de función de correlación en (3.82) es Toeplitz y Hermitiana, debido a  $r_x[-k] = r_x^*[k]$ , por lo que la obtención de los valores  $\sigma^2, a_1, a_2, \dots, a_p$  se puede hacer empleando el algoritmo de *Levinson-Durbin* [29].

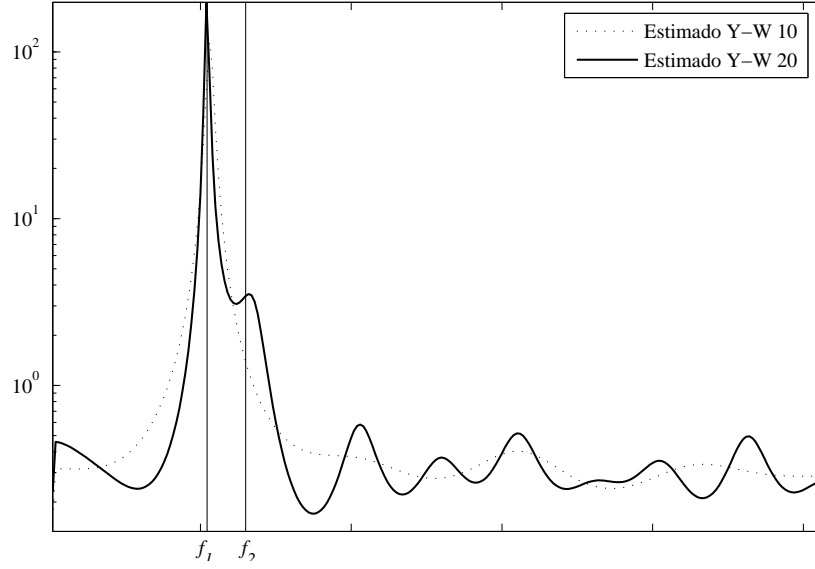
La relación de los valores, entre la función de correlación propia y los parámetros del modelo MA, se puede obtener de (3.80) asumiendo  $p = 0$  y teniendo en cuenta que para los sistemas no recursivos se cumple que  $h[k] = b_k, \forall 1 \leq k \leq q$ , entonces,

$$r_x[k] = \begin{cases} 0, & k > q; \\ \sigma^2 \sum_{l=k}^q b_l b_{l-k}^*, & 0 \leq k \leq q \\ 0, & k < 0 \end{cases} \quad (3.83)$$

Luego, la relación entre los parámetros MA y la función de correlación tiene carácter no lineal, condicionado por la convolución en (3.83). Otra forma conocida de estimar los parámetros del modelo AR, consiste en hallar directamente la solución del sistema de ecuaciones lineales, que se obtiene de la relación (3.80) para los  $p$  valores del índice  $k$ , ubicados en el intervalo  $q \leq k \leq q+p-1$ . Los valores de los parámetros AR se encuentran

de la solución del siguiente sistema de ecuaciones:

$$\begin{bmatrix} r_x[q] & r_x[q-1] & \cdots & r_x[q-p+1] \\ r_x[q+1] & r_x[q] & \cdots & r_x[q-p+2] \\ \vdots & \vdots & \ddots & \vdots \\ r_x[q+p-1] & r_x[q+p-2] & \cdots & r_x[q] \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_p \end{bmatrix} = \begin{bmatrix} r_x[1] \\ r_x[2] \\ \vdots \\ r_x[p] \end{bmatrix} \quad (3.84)$$



**Figura 3.10.** Ejemplo 3.10 calculado mediante el sistema modificado de Yule-Walker

La expresión (3.84) se denomina sistema de *ecuaciones normales de Yule-Walker* para los modelos ARMA o *sistema modificado de Yule-Walker*.

En la Figura 3.10, se muestra el ejemplo 3.10, calculado mediante el sistema modificado de Yule-Walker para diferentes órdenes del modelo.

## Problemas

**Problema 3.16.** Demostrar la igualdad de la ecuación (3.79).

**Problema 3.17.** Implementar el algoritmo de Levinson-Durbin para las ecuaciones normales de Yule-Walker para un proceso AR.

**Problema 3.18.** Determinar la media y la autocorrelación de la secuencia  $x[n]$ , que es salida de un proceso MA descrito por la ecuación en diferencias

$$x[n] = w[n] - 2w[n-1] + w[n-2]$$

donde  $w[n]$  es RBG con varianza  $\sigma_w^2$ .

**Problema 3.19.** Considerar un proceso MA descrito por la ecuación en diferencias

$$x[n] = w[n] + 0.81w[n-2]$$

donde  $w[n]$  es un RBG con varianza  $\sigma_w^2$ . Determinar los parámetros de los modelos AR de órdenes  $p = 2, 4, 8$ .

**Problema 3.20.** Hallar el correlograma y la DEP de la estimación del valor medio de longitud  $q$ :

$$x[n] = \frac{1}{q} (u[n] + u[n-1] \cdots + u[n-q+1])$$

**Problema 3.21.** Hallar la DEP, expresada mediante la respectiva función de correlación propia, para los procesos generados mediante un filtro formante con las siguientes funciones de transferencia

$$(a). H(z) = b_0 + b_1 z^{-1}. \quad (b). H(z) = b_0 + b_1 z^{-1} + b_2 z^{-2}.$$

**Problema 3.22.** Hallar el valor medio de la sucesión aleatoria descrita por la siguiente ecuación de diferencias:

$$x[n] = \alpha x[n-1] + u[n], \quad n \geq 0, \quad x[-1] = 0$$

donde  $u[n]$  es RBG discreto con media  $m_u$  y función de correlación  $r_u = \sigma_u^2 \delta[k]$

# Capítulo 4

## Procesos aleatorios de Markov

### 4.1. Definición y clasificación

Un proceso aleatorio  $\xi(t)$  se denomina *proceso de Markov* cuando para un valor fijo de  $\xi(u)$ , los demás valores aleatorios de  $\xi(t)$ ,  $t > u$ , no dependen de  $\xi(s)$ ,  $s < u$  [15]. Esto es, si para cualquier conjunto de argumentos de tiempo  $t_1, t_2, \dots, t_n$  definidos en el intervalo de análisis  $[0, T]$  la FDP condicional del último (extremo) valor  $\xi(t_n)$ , con  $\xi(t_1), \xi(t_2), \dots, \xi(t_{n-1})$ , depende solamente de  $\xi(t_{n-1})$ , con lo cual, para el conjunto  $\xi_1, \xi_2, \dots, \xi_n$ , es cierta la siguiente relación:

$$P(\xi(t_n) \leq \xi_n | \xi(t_1) = \xi_1, \dots, \xi(t_{n-1}) = \xi_{n-1}) = P(\xi(t_n) \leq \xi_n | \xi(t_{n-1}) = \xi_{n-1})$$

Una forma alterna de descripción del proceso de Markov está en su definición simétrica en los momentos de tiempo  $t_m, t_n, t_l$ :

$$\begin{aligned} P(\xi(t_m) \leq \xi_m, \xi(t_l) \leq \xi_l | \xi(t_n) = \xi_n) \\ = P(\xi(t_m) \leq \xi_m | \xi(t_n) = \xi_n) P(\xi(t_l) \leq \xi_l | \xi(t_n) = \xi_n) \end{aligned} \quad (4.1)$$

lo cual implica, que para un estado determinado del proceso, en el momento actual de tiempo  $t_n$ , los estados futuro  $t_l$  y pasado  $t_m$  cumplen la condición de independencia estadística.

De las anteriores definiciones, para los procesos de Markov, se deduce que la FDP con dimensión  $n$  que brinde la mayor información se puede representar en la forma,

$$p(\xi_1, \xi_2, \dots, \xi_n) = p(\xi_1) \prod_{i=1}^{n-1} p(\xi_{i+1} | \xi_i) \quad (4.2)$$

esto es, cualquier FDP con dimensión  $n$  de un proceso de Markov, se puede hallar empleando (4.2), cuando se conocen la FDP con dimensión simple del proceso y las densidades de probabilidad (o probabilidades) condicionales de cambio o *transición*. En forma general, la evolución de las probabilidades de cambio,  $P(\xi(t) \leq \xi | \xi(t_0) = \xi_0)$ , se describe por la

ecuación:

$$\frac{d}{dt}P = \mathcal{K}\{P\} \quad (4.3)$$

donde  $\mathcal{K}$  es un cierto operador lineal (matriz, diferenciador, etc.). La descripción operacional del sistema en (4.3), permite analizar el comportamiento de los procesos de Markov, empleando métodos conocidos de solución para las respectivas ecuaciones diferenciales, en las cuales el carácter de las tareas físicas desarrolladas pueden ser variadas, de acuerdo a las condiciones dadas iniciales y límite del sistema.

Sea el comportamiento de un sistema que se describe por alguna ecuación que cumpla (4.1), para unas condiciones iniciales dadas (el estado del sistema en el momento inicial  $t_0$ ). Si no se asume alguna otra restricción se debe hallar directamente la solución de (4.3), en caso contrario las condiciones límite deben reflejar las restricciones de solución. Las mismas restricciones tienen diferente naturaleza. Por ejemplo, un punto que refleja el comportamiento del sistema, que comienza su movimiento desde el estado inicial del sistema  $\xi(t_0) = \xi_0$ , cuando llegue al límite  $\xi(t) = c$  puede absorberse y el funcionamiento del sistema se detiene (Figura 4.1, trayectoria 1).

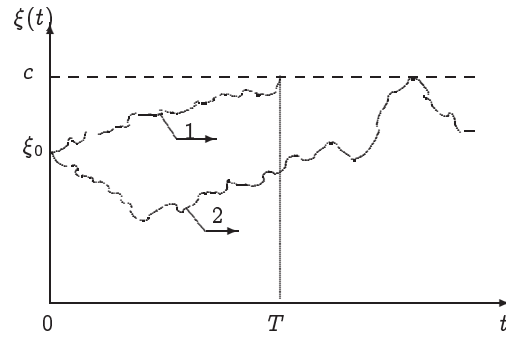
En otro caso, el punto puede reflejarse de la frontera  $c$  (trayectoria 2), y así sucesivamente. Si la frontera  $c$  es absorbente, entonces además de estimar la probabilidad de transición  $P$  para  $\xi < c$ , se puede hallar la probabilidad de absorción en algún tiempo  $T$ , o en la esperanza matemática, varianza o algún otro momento de tiempo  $T$ , en el cual el sistema por primera vez alcanza la frontera  $c$ .

La ecuación (4.3) tiene las siguientes formas de caracterización de los procesos de Markov:

1. Procesos de tiempo discreto y espacio discreto de fase,
2. Procesos de tiempo discreto y espacio continuo de fase
3. Procesos de tiempo continuo y espacio discreto de fase,
4. Procesos de tiempo continuo y espacio continuo de fase.

#### 4.1.1. Cadenas de Markov

Sea una señal aleatoria  $\theta(t)$  que toma una cantidad finita  $K$  de valores discretos, pertenecientes al conjunto  $\{\vartheta_1, \dots, \vartheta_K\}$ , tal que en los momentos determinados del tiempo ( $t_0 < t_1 < t_2 < \dots$ ) los valores del proceso aleatorio cambian de manera abrupta (con probabilidades de



**Figura 4.1.** Influencia de la frontera en la trayectoria del proceso

cambio conocidas), esto es, los cambios forman la sucesión aleatoria  $\theta_0 \rightarrow \theta_1 \rightarrow \theta_2, \dots$ , siendo  $\theta_n = \theta(t_n)$  el valor de la sucesión para el intervalo  $n$  de tiempo.

Una *cadena compuesta de Markov* de orden  $m \in \mathbb{Z}$  implica que la probabilidad de un nuevo valor del proceso depende, sola y estrictamente, de los  $m$  valores que lo anteceden:

$$P(\theta_n | \theta_0, \theta_1, \dots, \theta_{n-1}) = P(\theta_n | \theta_{n-m}, \dots, \theta_{n-1})$$

En general, una cadena de orden  $m$  se puede representar por una cadena simple ( $m = 1$ ), razón por la cual, de forma usual, el análisis de los procesos de Markov se realiza sobre cadenas simples, para las cuales las probabilidades conjuntas y finitas, determinadas a partir de (4.2), se expresan por medio de las *probabilidades de cambio*  $P(\theta_\mu | \theta_{\mu-1})$ :

$$P(\theta_0, \theta_1, \dots, \theta_n) = P(\theta_0) \prod_{\mu=1}^n P(\theta_\mu | \theta_{\mu-1})$$

Las cadenas simples de Markov, en la mayoría de los casos prácticos, se usan para la estimación de los diferentes valores de probabilidad en los momentos del tiempo  $t_n > t_0$ , y particularmente, cuando  $n \rightarrow \infty$ . En la descripción de las cadenas, es usual el empleo de las siguientes notaciones para los vectores columna de la probabilidades marginales y matrices de las probabilidades condicionales (ambas probabilidades deben cumplir con las condiciones de realización física y normalización):

$$\left\{ \begin{array}{ll} \mathbf{P}(n) &= \{p_k(n) = P(\theta_n = \vartheta_k) : k = 1, \dots, K\}, \quad n = 0, \dots, N \\ \mathbf{\Pi}(\mu, n) &= \{\pi_{mk}(\mu, n) = P(\theta_n = \vartheta_k | \theta_\mu = \vartheta_m) : m, k = 1, \dots, K\}, \quad 0 \leq \mu \leq n, n = 0, \\ p_k(n) &\geq 0, \quad \sum_{k=1}^K p_k(n) = 1, \quad n = 0, \dots, N \\ \pi_{mk}(\mu, n) &\geq 0, \quad \sum_{k=1}^K \pi_{mk}(\mu, n) = 1, \quad m = 1, \dots, K \end{array} \right. \quad (4.4)$$

El valor de  $p_k(n)$  es el valor de la probabilidad marginal de  $\vartheta_k$  en el instante  $n$ , cuando  $t = t_n$ . Mientras,  $\pi_{mk}(\mu, n)$  determina la probabilidad condicional del valor  $\vartheta_k$  en  $t_n$ , si en el momento anterior  $t_\mu < t_n$  el valor del proceso fue igual a  $\vartheta_m$ . Basados en la relación de probabilidad completa (2.3), para el sistema de (4.4), se escribe la ecuación de Markov:

$$\begin{aligned} \mathbf{P}(n) &= \mathbf{\Pi}^\top(\mu, n) \mathbf{P}(\mu) \\ \mathbf{\Pi}(\mu, n) &= \mathbf{\Pi}(\mu, m) \mathbf{\Pi}(m, n) = \prod_{i=0}^{n-\mu-1} \mathbf{\Pi}(\mu + i, \mu + i + 1), \quad n > m > \mu \geq 0 \end{aligned} \quad (4.5)$$

con lo cual, para la determinación de la matriz  $\Pi(\mu, n)$ ,  $\mu \leq n$ , es suficiente conocer la sucesión de matrices por cada intervalo de las probabilidades de cambio. Más aún, se puede afirmar que la descripción completa de una cadena simple de Markov se alcanza con la probabilidad del estado inicial y la matriz de las probabilidades de transición de la sucesión:

$$\Pi = \begin{bmatrix} \pi_{11} & \pi_{12} & \cdots & \pi_{1K} \\ \pi_{21} & \pi_{22} & \cdots & \pi_{2K} \\ \cdots & \cdots & \cdots & \cdots \\ \pi_{K1} & \pi_{K2} & \cdots & \pi_{KK} \end{bmatrix} \quad (4.6)$$

En las cadenas *homogéneas*, la probabilidad de cambio es invariable al tiempo y depende solamente de la diferencia de argumentos:

$$\Pi(\mu, n) = \Pi(n - \mu), \quad n > \mu \geq 0$$

A partir de (4.5) y al notar  $\Pi = \Pi(1)$ , se obtiene que la matriz de transición en una cadena simple y homogénea de Markov, después de  $n$  pasos, es igual a:

$$\Pi(n) = \Pi^n, \quad \mathbf{P}^\top(n) = \mathbf{P}^\top(0) \Pi^n \quad (4.7)$$

**Ejemplo 4.1.** Hallar la probabilidad de que después de  $n$  pasos la variable aleatoria pasa del estado  $\vartheta_i$  al estado  $\vartheta_j$ , si el proceso corresponde a una sucesión simple homogénea de Markov con  $K$  estados.

La matriz de transición corresponde a la expresión (4.6), de la cual se observa que la probabilidad de que la variable aleatoria pase del estado  $\vartheta_i \rightarrow \vartheta_j$  en un solo paso es igual a  $\pi_{ij}$ . Al punto  $j$  en dos pasos, la sucesión aleatoria puede llegar de dos maneras:  $i \rightarrow l; l \rightarrow j$ . La probabilidad de que la sucesión tome exactamente esa trayectoria es igual a  $\pi_{il}\pi_{lj}$ . Sin embargo, la cantidad de posibles puntos es  $K$ , por lo que teniendo en cuenta el teorema (2.1), se obtiene

$$\pi_{ij}(2) = \sum_{l=1}^K \pi_{il}\pi_{lj}, \quad \Pi(2) = \{\pi_{ij}(2) : i, j = 1, \dots, K\}$$

La sumatoria abierta del elemento  $\pi_{ij}(2)$  es igual a,

$$\pi_{ij}(2) = \sum_{l=1}^K \pi_{il}\pi_{lj} = \pi_{i1}\pi_{1j} + \pi_{i2}\pi_{2j} + \cdots + \pi_{iK}\pi_{Kj}$$

la cual corresponde a la expresión para la multiplicación de matrices, esto es,

$$\Pi(2) = \Pi(1) \Pi(1) = \Pi^2$$

De manera similar, se obtiene la matriz para  $3, 4, \dots, n$  pasos:  $\Pi(n) = \Pi^n$ .



**Ejemplo 4.2.** Sea la matriz de transición

$$\mathbf{\Pi} = \begin{bmatrix} 1/2 & 1/2 \\ 2/5 & 3/5 \end{bmatrix}$$

Hallar la matriz de transición para dos pasos y calcular las probabilidades de que el sistema llega al estado 1 en un paso, además al estado 2 en dos pasos, si se conoce el estado inicial:  $p_0(1) = 1$  y  $p_0(2) = 0$ .

Por cuanto,  $\mathbf{\Pi}(2) = \mathbf{\Pi}^2$ , entonces

$$\mathbf{\Pi}(2) = \begin{bmatrix} 1/2 & 1/2 \\ 2/5 & 3/5 \end{bmatrix} \begin{bmatrix} 1/2 & 1/2 \\ 2/5 & 3/5 \end{bmatrix} = \begin{bmatrix} \pi_{11}(2) & \pi_{12}(2) \\ \pi_{21}(2) & \pi_{22}(2) \end{bmatrix} = \begin{bmatrix} 9/20 & 11/20 \\ 11/25 & 14/25 \end{bmatrix}$$

las probabilidades absolutas  $p_j(n)$ , teniendo en cuenta el vector dado inicial  $\mathbf{P}^T(0) = [1 \ 0]$ , se hallan de la expresión (4.7),

$$\begin{aligned} \mathbf{P}^T(1) &= [1 \ 0] \begin{bmatrix} 1/2 & 1/2 \\ 2/5 & 3/5 \end{bmatrix} = [1/2 \ 1/2] \\ \mathbf{P}^T(2) &= [1 \ 0] \begin{bmatrix} 9/20 & 11/20 \\ 11/25 & 14/25 \end{bmatrix} = [9/20 \ 11/20] \end{aligned}$$

La lectura de los resultados obtenidos es la siguiente: si la sucesión inicialmente se encontraba en el primer estado, entonces la probabilidad de que en un paso la trayectoria se quede en este mismo valor de estado es igual a  $1/2$ , e igual a  $1/2$ , si toma el segundo estado. Para estas mismas condiciones iniciales, la probabilidad de que en dos pasos la sucesión tome el segundo estado es igual a  $9/20$ , mientras la probabilidad de tomar el segundo estado es  $9/20$ .

Una cadena homogénea se denomina *estacionaria* si se cumple que  $\mathbf{P}(n) = \mathbf{P}$ , lo cual implica que no depende del valor normalizado del tiempo  $n$ . En forma general, las probabilidades *finales*  $\mathbf{P}$ , si existen, se hallan como resultado del límite de la matriz de transición,

$$\mathbf{P} = \lim_{n \rightarrow \infty} \mathbf{P}(n) \quad (4.8)$$

Sin embargo, si se tiene que las probabilidades iniciales  $\mathbf{P}(0)$  coinciden con las probabilidades finales correspondientes  $\mathbf{P}$ , entonces la cadena de Markov es estacionaria a partir del momento de tiempo  $t_0$ .

Las probabilidades finales deben satisfacer el siguiente sistema lineal de ecuaciones algebraicas de orden  $K$ :

$$(\mathbf{I} - \mathbf{\Pi}^T) \mathbf{P} = \mathbf{0} \quad (4.9)$$

además de la condición adicional  $\sum_{k=1}^K p_k = 1, p_k \geq 0$ , siendo  $\mathbf{I}$  la matriz unitaria. Debido a la condición (4.4) las  $K$  ecuaciones de (4.9) son linealmente dependientes. Por lo tanto, las  $K$  probabilidades finales se deben determinar a partir de las ecuaciones de (4.9).

La clasificación de los estados de un proceso de Markov depende de si, dado un estado, el proceso puede llegar hasta cualquier otro estado. Un estado  $\vartheta_j$  se denomina *sin retorno*

si existe un estado  $\mathfrak{v}_k, k \neq i$ , después de una cantidad de pasos  $n$ , tales que  $\pi_{jk}(n) > 0$ , pero  $\pi_{kj} = 0, \forall m$ . Los estados *con retorno* suponen el retorno desde cualquier estado del proceso, inclusive si la cantidad de pasos en la dirección de ida y en la de retorno no son iguales. De otra parte, la pareja de estados se denomina *conectiva*, si dados  $n, m$  se cumplen las condiciones  $\pi_{jk}(n) > 0$  y  $\pi_{kj}(m) > 0$ . El conjunto de estados con retorno y conectivos se denomina *ergódico*. Por cierto, si dado un valor suficientemente grande de pasos  $n_0$ , se cumple que todos los elementos de la matriz  $\mathbf{\Pi}^n$  son todos positivos para cualquier  $n > n_0$ , entonces el proceso se define como *regularmente ergódico*.

En los procesos ergódicos, se define la matriz fundamental, por medio de la cual se pueden hallar otras características estadísticas básicas del proceso, en la forma

$$\mathbf{Z} = (\mathbf{I} + \mathbf{\Pi} + \mathbf{G})^{-1}$$

donde  $\mathbf{G}$  es la matriz, en la cual cada fila es igual al vector  $\mathbf{P}^T$  de las probabilidades finales en (4.8).

**Ejemplo 4.3.** Hallar el valor asintótico de las probabilidades absolutas  $\mathbf{P}(n)$ , para  $n \rightarrow \infty$ , de una cadena ergódica de Markov.

La ecuación (4.7) determina las probabilidades absolutas después de  $n$  pasos de la sucesión, para la cual se analiza la siguiente relación límite por cada uno de sus términos elementales,

$$\lim_{n \rightarrow \infty} p_j(n) = \lim_{n \rightarrow \infty} \sum_{i=1}^K p_i(0) \pi_{ij}(n) = \sum_{i=1}^K p_i(0) \lim_{n \rightarrow \infty} \pi_{ij}(n)$$

Al tener en cuenta, los valores finales de probabilidad, en correspondencia con la definición de ergodicidad, se obtiene,

$$\lim_{n \rightarrow \infty} p_j(n) = \sum_{i=1}^K p_i(0) \pi_{ij} = \pi_{ij} \sum_{i=1}^K p_i(0)$$

que muestra la validez de (4.8), debido a que  $\sum_{i=1}^K p_i(0) = 1$ .

**Ejemplo 4.4.** Establecer si la matriz de transición del ejemplo 4.2 es o no ergódica; en tal caso, calcular las probabilidades finales.

En el ejemplo 4.2 se obtuvo la matriz de transición para 2 pasos igual a  $\mathbf{\Pi}(2) = \mathbf{\Pi}^2$ , con todos su elementos positivos,  $\{\pi_{ij}(2) \geq 0\}$ . De lo cual se puede afirmar que los elementos de la matriz  $\mathbf{\Pi}^n$  para  $n \rightarrow \infty$ , también serán estrictamente positivos y, por lo tanto, la cadena es ergódica.

En cuanto a las probabilidades finales, éstas se calculan resolviendo las condiciones en (4.9), que para cada uno de los elementos toma la forma,

$$\begin{cases} p_i &= \sum_{l=1}^K p_l \pi_{li} \\ \sum_{i=1}^K p_i &= 1, p_i > 0 \end{cases}$$

En el caso concreto, se tiene las siguientes ecuaciones:

$$\begin{cases} p_1 = p_1 \pi_{11} + p_2 \pi_{21} = p_1 1/2 + p_2 2/5 \\ p_2 = p_1 \pi_{12} + p_2 \pi_{22} = p_1 1/2 + p_2 3/5 \\ p_1 + p_2 = 1 \end{cases}$$

De la primera ecuación se tiene  $p_1 1/2 = p_2 2/5$ ,  $p_1 1/2 = p_2 2/5$ ,  $p_1 = 1 - p_2$ , entonces,  $p_1 = 4/9$ ,  $p_2 = 5/9$ . Los resultados obtenidos al resolver el problema 4.1, en el primer caso, muestran que los elementos de las matrices  $\Pi^3, \Pi^4, \dots$ , convergen a los resultados obtenidos.

#### 4.1.2. Procesos discretos de Markov

Sea un proceso aleatorio  $\theta(t)$  que toma solamente valores discretos  $\{\vartheta_k : k = 1, \dots, K\}$ , pero cuyo cambio no ocurre en momentos fijos, sino en valores aleatorios del tiempo (Figura 4.2), caso en el cual las notaciones de (4.4) se formulan de nuevo así:

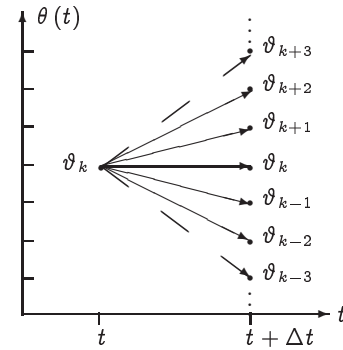
$$\begin{cases} \mathbf{P}(t) &= \{p_k(t) = P(\theta(t) = \vartheta_k) : k = 1, \dots, K\} \\ \Pi(t_0, t) &= \{\pi_{mk}(t_0, t) = P(\theta(t) = \vartheta_k \mid \theta(t_0) = \vartheta_m) : m, k = 1, \dots, K\}, \quad 0 \leq t_0 \leq t \\ p_k(t) &\geq 0, \quad \sum_{k=1}^K p_k(t) = 1 \\ \pi_{mk}(t_0, t) &\geq 0, \quad \sum_{k=1}^K \pi_{mk}(t_0, t) = 1, \quad j = 1, \dots, K \end{cases} \quad (4.10)$$

Similar a (4.5), para las probabilidades de cambio se cumple la *ecuación de Chapman-Kolmogorov*:

$$\Pi(t_0, t + \Delta t) = \Pi(t_0, t) \Pi(t, t + \Delta t), \quad t > t_0, \quad \Delta t > 0 \quad (4.11)$$

En los procesos discretos de Markov se asume la propiedad de *ordinariedad* [15], según la cual, para intervalos relativamente pequeños de tiempo  $\Delta t$  entre los discretos, se considera que la probabilidad  $\pi_{kk}$  de que el valor actual no cambie, excede la probabilidad de transición de este valor, esto es,

$$\begin{aligned} \pi_{kk}(t, t + \Delta t) &= 1 + a_{kk}(t) \Delta t + o(\Delta t) \\ \pi_{mk}(t, t + \Delta t) &= a_{kl}(t) \Delta t + o(\Delta t) \end{aligned} \quad (4.12)$$



**Figura 4.2.** Procesos discretos

donde  $o(\Delta t)$  denota los términos residuales, considerados insignificantes, mayores al de primer orden con relación a  $\Delta t$ :  $\lim_{\Delta t \rightarrow 0} \frac{o(\Delta t)}{\Delta t} = 0$ .

De acuerdo con la condición de estacionariedad (4.7), de la última expresión se tiene que

$$a_{kk}(t) = - \sum_{m \neq k} a_{km}(t) \leq 0, \quad a_{km}(t) \geq 0 \quad (4.13)$$

Además, tienen lugar las siguientes igualdades:

$$\Pi(t_0, t_0) = \mathbf{I}, \quad \Pi(t_0, t + \Delta t) = \mathbf{I} + \mathbf{A}(t) \Delta t + \mathbf{o}(\Delta t) \quad (4.14)$$

donde  $\mathbf{A}(t)$  corresponde a la matriz *infinitesimal de probabilidades*. Reemplazando la igualdad (4.14) en (4.11), y haciendo  $\Delta t \rightarrow 0$  se obtiene la *ecuación de Kolmogorov en el sentido directo* (esto es, orientada en el sentido natural de desarrollo del tiempo):

$$\frac{\partial}{\partial t} \Pi(t_0, t) = \Pi(t_0, t) \mathbf{A}(t) \quad (4.15)$$

cuya solución general, con condiciones iniciales (4.14) y cuando  $\mathbf{A}(t) = \mathbf{A} = \text{const.}$ , corresponde a la matriz exponencial

$$\Pi(t_0, t) = \exp(\mathbf{A}(t - t_0)) \quad (4.16)$$

Un proceso discreto de Markov conserva su tipo, cuando se analiza su desarrollo en el sentido contrario del tiempo, para el que se tiene la *ecuación inversa de Kolmogorov*:

$$\frac{\partial}{\partial t_0} \Pi(t_0, t) = -\mathbf{A}(t_0) \Pi(t_0, t), \quad t \geq t_0$$

que tiene como solución, asumiendo  $\mathbf{A}(t) = \mathbf{A} = \text{const.}$ , la matriz de forma exponencial

$$\mathbf{P}(t_0, t) = \exp(\mathbf{A}^\top(t - t_0)) \mathbf{P}(t_0)$$

A su vez, un proceso discreto de Markov se denomina *homogéneo* si la matriz de transición depende solamente de la diferencia  $\tau = t - t_0$ , entonces:

$$\Pi(t_0, t) = \Pi(t - t_0) = \Pi(\tau)$$

De la condición (4.12), se observa que para un proceso discreto de Markov homogéneo, la matriz de transición de probabilidades infinitesimales  $\mathbf{A}(t) = \mathbf{A}$  no depende del tiempo y el sistema de ecuaciones diferenciales de (4.15) se simplifica hasta la expresión:

$$\frac{\partial}{\partial \tau} \Pi(\tau) = \Pi(\tau) \mathbf{A} \quad (4.17)$$

Un proceso discreto de Markov se considera *ergódico*, si para  $\tau \rightarrow \infty$  existen los valores límites de las probabilidades de estados  $\mathbf{P} = \lim_{\tau \rightarrow \infty} \mathbf{P}(\tau)$ . Así mismo, dada la condición  $\sum_{k=1}^K p_k = 1$ , las probabilidades de los valores estacionarios se determinan del sistema algebraico de ecuaciones con  $K - 1$  variables:  $\mathbf{A}^\top \mathbf{P} = \mathbf{0}$ .

**Ejemplo 4.5.** Hallar la probabilidad  $P_k(t)$  de que un proceso discreto homogéneo de Markov en el momento de tiempo  $t$  se encuentre en el estado  $\vartheta_k$ ,  $k = 1, \dots, K$ .

La solución básicamente consiste en hallar la probabilidad  $P_k(t + \Delta t)$  de que el proceso tome el valor  $\vartheta_k$  en un tiempo  $(t + \Delta t)$ , para luego hallar la relación límite de cambio. En este caso, se considera que el proceso puede llegar  $\vartheta_k$ , con probabilidad de cambio  $\pi_{lk}(t, t + \Delta t)$  a partir del estado  $\vartheta_l$  en los siguientes casos:

1.  $\vartheta_k \rightarrow \vartheta_k$ , el proceso no cambia de estado con probabilidad de transición  $\pi_{kk}(t, t + \Delta t)$ .
2.  $\vartheta_{k-1} \rightarrow \vartheta_k \rightarrow \pi_{k-1,k}(t, t + \Delta t)$
3.  $\vartheta_{k+1} \rightarrow \vartheta_k \rightarrow \pi_{k+1,k}(t, t + \Delta t)$
4.  $\vartheta_{k \pm q} \rightarrow \vartheta_k \rightarrow \pi_{k \pm q,k}(t, t + \Delta t)$ , para  $q \geq 2$ , cuando el sistema varia, bien a niveles superiores o bien sea a niveles inferiores, en dos o más estados.

Las respectivas probabilidades de transición se calculan de la ecuación (4.12) que, teniendo en cuenta la homogeneidad del proceso, toma la forma

$$\begin{cases} \pi_{k,k+1}(\Delta t) &= \alpha_k \Delta t + o(\Delta t), \alpha \geq 0 \\ \pi_{k,k-1}(\Delta t) &= \beta_k \Delta t + o(\Delta t), \beta \geq 0, \alpha\beta > 0 \\ \pi_{k,\pm q}(\Delta t) &= o(\Delta t), q \geq 2 \end{cases}$$

con lo cual se tiene,

$$\begin{aligned} \pi_{k,k}(\Delta t) &= P_k(t)(1 - \pi_{k,k+1}(\Delta t) - \pi_{k,k-1}(\Delta t) - \pi_{k,\pm q}(\Delta t)) \\ &= P_k(t)(1 - \alpha_k \Delta t - \beta_k \Delta t - 3o(\Delta t)) \\ \pi_{k,k-1}(\Delta t) &= P_{k-1}(t)(\beta_{k-1} \Delta t + o(\Delta t)) \\ \pi_{k,k+1}(\Delta t) &= P_{k+1}(t)(\alpha_{k+1} \Delta t + o(\Delta t)) \\ \pi_{k,k \pm q}(\Delta t) &= P_{k \pm q}(t)o(\Delta t) \end{aligned}$$

Por cuanto, los cuatro eventos son independientes y no excluyentes, entonces, la probabilidad  $P_k(t + \Delta t)$  es igual a la suma de las anteriores cuatro probabilidades de transición,

$$P_k(t + \Delta t) = P_k(t) - (\alpha_k + \beta_k) P_k(t) \Delta t + \alpha_{k-1} P_{k-1}(t) \Delta t + \beta_{k+1} P_{k+1}(t) \Delta t + O_k(\Delta t) \quad (1)$$

donde  $O_k(\Delta t) = -3P_k(t)o(\Delta t) + P_{k-1}(t)o(\Delta t) + P_{k+1}(t)o(\Delta t) + P_{k \pm q}(t)o(\Delta t)$  es el término relacionado con la probabilidad de transición del estado actual a otro estado que se diferencie en dos o más niveles, y la cual se considera infinitamente pequeña, esto es, se asume el cambio suficientemente lento del proceso, tal que no se consideran saltos de más de un nivel de diferencia.

De la ecuación (1), asumiendo  $O_k(\Delta t) = 0$ , entonces se tiene que

$$\frac{P_k(t + \Delta t) - P_k(t)}{\Delta t} = -(\alpha_k + \beta_k) P_k(t) + \alpha_{k-1} P_{k-1}(t) + \beta_{k+1} P_{k+1}(t)$$

que al realizar el cambio límite da como resultado la expresión,

$$\frac{dP_n(t)}{dt} = -(\alpha_k + \beta_k) P_k(t) + \alpha_{k-1} P_{k-1}(t) + \beta_{k+1} P_{k+1}(t) \quad (2)$$

De forma similar se obtiene la relación para probabilidad del estado más bajo del sistema  $\vartheta_1$ , asumiendo  $P_{k-1}(t) = 0$  y  $\beta_1 = 0$ , mientras para el estado más alto,  $\vartheta_K$ , se asume,  $P_{k+1}(t) = 0$  y  $\alpha_K = 0$ . La generalización de (2) corresponde a la ecuación (4.17) para un proceso discreto homogéneo con  $K$  estados.

Cabe anotar que la forma de cambio de los parámetros  $\alpha_k = \alpha(k)$ ,  $\beta_k = \beta(k)$  determina la dependencia en el tiempo de  $P_k(t)$  (o  $P_k$  en los procesos ergódicos). Así por ejemplo, si

se tiene  $\beta_k = 0$  y  $\alpha_k = \lambda$ , entonces la probabilidad  $P_k(t)$  corresponde a la FDP de Poisson con parámetro  $\lambda$ , mostrada en el ejemplo 2.8. En cambio, cuando  $\alpha_k = \alpha$ , y  $\beta_k = k\beta$ , asumiendo la ergodicidad del proceso, la probabilidad  $P_k(t)$  corresponde a la FDP de Erlang, que corresponde a un caso particular de la  $\Gamma$ -densidad (2.28).

Cuando uno de los parámetros correspondientes a las probabilidades de transición, presentes en la ecuación (1), se asume igual a cero, implica que el proceso tiene una sola dirección de desarrollo: si  $\alpha(k) = 0, \forall k$ , entonces, el proceso nunca crece y su desarrollo corresponde a la disminución del orden del estado (desintegración o muerte), cuando  $\beta(k) = 0, \forall k$ , el proceso nunca decrece y aumenta de forma progresiva el orden de los estados (nacimiento).

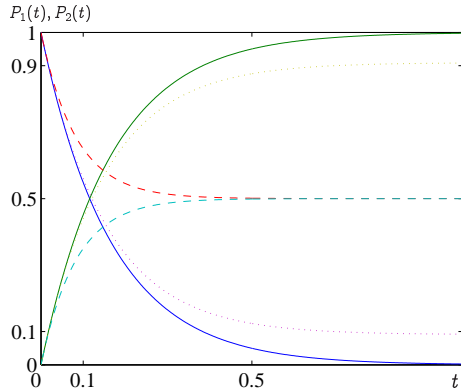
**Ejemplo 4.6.** Un proceso discreto homogéneo de Markov puede encontrarse en uno de dos posibles estados:  $\{\vartheta_k : k = 1, 2\}$ . El cambio del primer estado  $\vartheta_1 \rightarrow \vartheta_2$  ocurre con probabilidad  $\alpha dt$ , mientras el cambio  $\vartheta_2 \rightarrow \vartheta_1$  ocurre con probabilidad  $\beta dt$ . Hallar las respectivas probabilidades  $P_k(t)$  de que el proceso se encuentre en cada uno de los dos estados, asumiendo que el proceso en el tiempo  $t = 0$  se encuentra en el estado  $\vartheta_1$ .

De las condiciones del proceso se tiene que  $A_{12} = \alpha$  y  $A_{21} = \beta$  y, teniendo en cuenta (4.13),  $A_{11} = -\alpha$ ,  $A_{22} = -\beta$ . Por lo tanto, la ecuación (4.15) para las probabilidades de estado definidas en (4.10)  $p_1(t) = P_1(t)$  y  $p_2(t) = P_2(t)$ , toma la forma

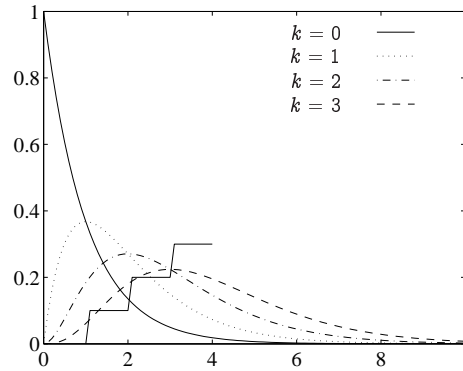
$$dP_1/dt = -\alpha P_1 + \beta P_2, \quad dP_2/dt = \alpha P_1 - \beta P_2 \quad (3)$$

Por cuanto,  $P_1 + P_2 = 1$ , se emplea una sola ecuación en (3), por ejemplo, la primera,

$$dP_1/dt = -\lambda P_1 + \beta, \quad \lambda = \alpha + \beta$$



(a) Proceso con dos estados



(b) Proceso de nacimiento y varios estados

**Figura 4.3.** Dependencia de las probabilidades de estado en el tiempo

Inicialmente, se asume el valor de  $P_1 = 1$ , para  $t = 0$ , esto es, al principio el proceso con certeza se encuentra en el estado  $\vartheta_1$ . Entonces, de (4.16) se obtiene la solución

$$P_1(t) = e^{-\lambda t} + \frac{\beta}{\lambda} (1 - e^{-\lambda t}), \quad P_2(t) = \frac{\alpha}{\lambda} (1 - e^{-\lambda t})$$

El análisis de ambas funciones de probabilidad, cuando  $t \rightarrow \infty$ , muestra que se obtienen los valores asintóticos constantes, que no dependen de las condiciones iniciales,

$$P_1 = \beta/\lambda, \quad P_2 = \alpha/\lambda$$

de lo cual se establece que el proceso es ergódico.

Como antes se dijo, los coeficientes  $\alpha$  y  $\beta$  determinan la forma de las probabilidades de estado del proceso, como se observa en la Figura 4.3(a), en la cual se muestran ejemplos de cálculo para  $\alpha = \beta$  (línea  $\dashrightarrow$ ),  $\alpha \gg \beta$  ( $\cdots$ ) y  $\beta = 0$  (con línea continua). Este último caso, que corresponde a un proceso de nacimiento,  $\vartheta_k \rightarrow \vartheta_{k+1}$ , se puede generalizar para  $\alpha(k) = \alpha$  y  $k$  estados,

$$P_k(t) = \frac{(\alpha t)^k}{k!} e^{-\alpha t}$$

Las respectivas dependencias para  $k = 0, 1, 2, 3$  se muestran en la Figura 4.3(b), en la cual además se muestra el estado más probable del procesos en cada momento del tiempo.

#### 4.1.3. Procesos continuos de Markov

El dominio del proceso aleatorio con valores continuos  $x(t)$  y el dominio de su definición  $[0, T]$  son conjuntos continuos, lo cual no significa que el proceso de Markov sea continuo, más bien, el término proceso de valores continuos significa que sus valores, definidos en un dominio continuo, cambian de manera abrupta en forma incremental en un dominio continuo de tiempo.

Sea la FDP  $p_\xi(x(t_k)|x(t_1), \dots, x(t_{k-1}))$  para la señal aleatoria  $\xi(t)$ , asumiendo que son conocidos todos los valores del proceso correspondientes a los momentos previos,  $t_1, \dots, t_{k-1}$ . La suposición Markoviana (4.2), en el caso continuo, se escribe en la forma,

$$p_\xi(x(t_k)|x(t_1), \dots, x(t_{k-1})) = p_\xi(x(t_k)|x(t_{k-1}))$$

En este caso, la FDP conjunta de los  $k$  valores considerados del proceso es igual a,

$$\begin{aligned} p_\xi(x(t_1), \dots, x(t_k)) &= p_\xi(x(t_k)|x(t_1), \dots, x(t_{k-1})) p_\xi(x(t_1), \dots, x(t_{k-1})) = \\ &= p_\xi(x(t_1)|x(t_{k-1})) p_\xi(x(t_{k-1})|x(t_{k-2})) \cdots \\ &\cdots p_\xi(x(t_2)|x(t_2)) p_\xi(x(t_1)) \end{aligned}$$

esto es, la FDP de los valores de un proceso de Markov de cualquier dimensión, se puede desdoblar a través de densidades de probabilidad de dos dimensiones del tipo  $p_\xi(x(t_n)|x(t_{n-1}))$  y la FDP  $p_\xi(t_1)$ , correspondiente al inicio del proceso.

De otra parte, teniendo en cuenta la relación unívoca entre las FDP condicionales y marginales del proceso, dadas por el teorema de Bayes 2.3, entonces se puede afirmar que todas las funciones de probabilidad de un proceso de Markov, de forma unívoca, también pueden expresarse mediante una FDP con dos dimensiones, correspondientes a dos valores cualesquiera del tiempo. A su vez, esta FDP de dos dimensiones caracteriza de manera completa el proceso.

Debido a la dinámica propia del proceso de Markov en el tiempo, es importante determinar la probabilidad de que la señal aleatoria  $\xi(t)$  con valor  $x_1$  para el momento de tiempo  $t_1$ , en el momento de tiempo  $t_3$  toma cualquier valor dentro del intervalo  $x_3 + \Delta x_3$ , si en

el momento de tiempo  $t_2$  el proceso tenía un valor dentro del intervalo  $x_2 + \Delta x_2$ , siendo  $t_1 < t_2 < t_3$ . Ambos eventos, les corresponden las siguientes dos probabilidades:

$$p(x(t_1)|x(t_2))dx_2$$

$$p(x(t_2)|x(t_3))dx_3$$

obtenidas a partir de la consideración de la condición de continuidad del proceso, y la cuales se consideran iguales dentro de los intervalos infinitamente pequeños,  $x_i + \Delta x_i$ ,  $i = 1, 2$ , de cambio del proceso.

Debido a que ambos sucesos analizados, para los procesos de Markov, se consideran independientes, entonces la probabilidad de cambio del proceso del estado  $x_1$  al estado  $x_3$ , con las condiciones de transitividad antes hechas, corresponde a la multiplicación de ambas probabilidades,

$$p(x(t_1)|x(t_2))dx_2 p(x(t_2)|x(t_3))dx_3 \quad (4.18)$$

De la última ecuación (4.23) se obtiene la probabilidad de cambio del estado  $x_1$  a  $x_3$ , teniendo en cuenta (2.39). Además, la misma probabilidad de cambio del estado  $x_1$  a  $x_3$  sin importar cual sea el valor del proceso en  $t_2$  (pero asumiendo aún la continuidad del proceso) corresponde a

$$p(x(t_1)|x(t_3))dx_3 \quad (4.19)$$

Igualando ambas expresiones (4.18) y (4.19) se obtiene la *ecuación generalizada de Markov*,

$$p(x(t_1)|x(t_3)) = \int_{-\infty}^{\infty} p(x(t_1)|x(t_2))p(x(t_2)|x(t_3))dx_2 \quad (4.20)$$

**Ejemplo 4.7.** Sea una partícula que en el momento  $k\Delta t$  salta al valor  $\pm\Delta x$ , con probabilidad  $p$  si aumenta ( $+\Delta x$ ) o con probabilidad  $q = 1 - p$  si disminuye ( $-\Delta x$ ). Basados en las relaciones obtenidas para los procesos discretos homogéneos de Markov, pero asumiendo su no estacionariedad debida al cambio en el tiempo de la distribución en el tiempo, hallar el respectivo sistema de ecuaciones diferenciales que describen el comportamiento estocástico de la partícula.

El cambio de los modelos de procesos discretos de Markov a las respectivas de los procesos continuos se realiza hallando el cambio límite de los valores  $\Delta \rightarrow 0$  y  $\Delta t \rightarrow 0$  para los modelos de ecuaciones diferenciales que los describen.

En particular, para las condiciones dadas de cambio de la partícula, la ecuación diferencial (4.15) tiene la forma,

$$P(t, x) = pP(t - \Delta t, x - \Delta t) + qP(t - \Delta t, x + \Delta t)$$

Se halla la descomposición por serie de potencias en términos de  $\Delta t$  y  $\Delta x$  de la parte derecha



de la ecuación:

$$\begin{aligned} P(t, x) = & pP(t, x) - p \frac{\partial P}{\partial t} \Delta t - p \frac{\partial P}{\partial x} \Delta x + qP(t, x) - q \frac{\partial P}{\partial t} \Delta t + \\ & + q \frac{\partial P}{\partial x} \Delta x + p \frac{\partial^2 P}{\partial t \partial x} \Delta t \Delta x + \frac{p}{2} \frac{\partial^2 P}{\partial t^2} \Delta t^2 + \frac{p}{2} \frac{\partial^2 P}{\partial x^2} \Delta x^2 - \\ & - q \frac{\partial^2 P}{\partial t \partial x} \Delta t \Delta x + \frac{q}{2} \frac{\partial^2 P}{\partial t^2} \Delta t^2 + \frac{q}{2} \frac{\partial^2 P}{\partial x^2} \Delta x^2 + \dots \end{aligned}$$

Debido a que  $p + q = 1$ , entonces se obtienen los primeros términos para la serie anterior:

$$0 = -\frac{\partial P}{\partial t} \Delta t - (p - q) \frac{\partial P}{\partial x} \Delta x + (p - q) \frac{\partial^2 P}{\partial t \partial x} \Delta t \Delta x + \frac{1}{2} \frac{\partial^2 P}{\partial t^2} \Delta t^2 + \frac{1}{2} \frac{\partial^2 P}{\partial x^2} \Delta x^2 + \dots$$

Luego, se dividen todos los términos de la serie de representación por  $\Delta t$ ,

$$\frac{\partial P}{\partial t} = -(p - q) \frac{\partial P}{\partial x} \frac{\Delta x}{\Delta t} + (p - q) \frac{\partial^2 P}{\partial t \partial x} \Delta x + \frac{1}{2} \frac{\partial^2 P}{\partial t^2} \Delta t + \frac{1}{2} \frac{\partial^2 P}{\partial x^2} \frac{\Delta x^2}{\Delta t} + \dots$$

Seguidamente, se halla el límite de la última expresión para  $\Delta t \rightarrow 0$ . Si se asume que la probabilidad de aparición de los valores grandes  $\Delta x$  decrece al disminuir el intervalo de tiempo  $\Delta t$ , tan rápido que todos los momentos de la diferencia  $\Delta x$ , comenzando desde el tercer discreto de tiempo tienden a cero más rápido que  $\Delta t$ .

De lo anterior, se obtiene los siguientes valores límite para cada uno de los términos de la serie:

1.  $\lim_{\Delta t \rightarrow 0} (p - q) \frac{\Delta x}{\Delta t} = a$
2.  $\lim_{\substack{\Delta t \rightarrow 0 \\ \Delta x \rightarrow 0 \\ (p-q) \rightarrow 0}} (p - q) \Delta x = 0$
3.  $\lim_{\Delta t \rightarrow 0} \frac{1}{2} \frac{\partial^2 P}{\partial t^2} \Delta t = 0,$
4.  $\lim_{\Delta t \rightarrow 0} \frac{\Delta x^2}{\Delta t} = b$

Por lo tanto, la operación de límite, reemplazando la probabilidad por la respectiva función densidad de probabilidad, resulta en la ecuación,

$$\frac{\partial p}{\partial t} = -a \frac{\partial p}{\partial x} + \frac{b}{2} \frac{\partial^2 p}{\partial x^2} \quad (1)$$

La solución de (1), teniendo en cuenta la condición de unidad de área,  $\int p(x, t) dx = 1$ , además, asumiendo las condiciones iniciales  $x_0$ , dadas en  $t_0$ , se demuestra que corresponde a la FDP Gaussiana,

$$p(x, t) = \frac{1}{\sqrt{2\pi b(t - t_0)}} \exp \left( -\frac{((x - x_0) - a(t - t_0))^2}{2b(t - t_0)} \right)$$

De la solución se observa que el coeficiente  $a$  corresponde al valor medio del tiempo o la velocidad de cambio del valor medio, mientras  $b$  es la varianza de los valores del tiempo.

En general, la densidad de probabilidad  $p(x(t_1)|x(t_3))$ , que cambia en función de los parámetros iniciales  $(x_1, t_1)$ , cumple la ecuación diferencial,

$$\frac{\partial p}{\partial t} + a(x, t) \frac{\partial p}{\partial x} + \frac{1}{2} b(t, x) \frac{\partial^2 p}{\partial t^2} = 0 \quad (4.21)$$

donde  $a(x, t)$  y  $b(t, x)$  pueden ser funciones variantes en el tiempo, que determinan las particularidades del proceso de Markov, y que corresponden a los siguientes momentos:

$$\begin{cases} a(x, t) &= \lim_{\Delta t \rightarrow 0} \frac{1}{\Delta t} \int_{-\infty}^{\infty} (x_2 - x_1) p(x(t_1)|x(t_1 + \Delta t)) dx_2 = \lim_{\Delta t \rightarrow 0} \frac{1}{\Delta t} E\{\Delta x|x\} \\ b(x, t) &= \lim_{\Delta t \rightarrow 0} \frac{1}{\Delta t} \int_{-\infty}^{\infty} (x_2 - x_1)^2 p(x(t_1)|x(t_1 + \Delta t)) dx_2 = \lim_{\Delta t \rightarrow 0} \frac{1}{\Delta t} E\{\Delta x^2|x\} \end{cases} \quad (4.22)$$

**Ejemplo 4.8.** Sea un proceso aleatorio al cual le corresponde el modelo en forma de la ecuación diferencial, que describe el movimiento Browniano,

$$\frac{d\xi}{dt} + \alpha \xi(t) = m\eta(t) \quad (1)$$

donde  $\alpha$  y  $m$  son constantes,  $\eta(t)$  es ruido blanco Gaussiano con valor medio igual a cero y función de correlación  $R_\xi(\tau) = \delta(\tau)$ . Hallar los momentos de la FDP del respectivo modelo de Markov.

Al asumir la total independencia de los valores del proceso  $\xi(t)$ , del modelo propuesto resulta que  $\eta(t)$  es la solución de una ecuación de primer orden, la cual se determina por las condiciones iniciales del proceso de manera unívoca. Además, los valores de la señal  $\eta(t)$  son independientes, cuyos valores pasados no influyen en los futuros. Por lo tanto, el proceso  $\xi(t)$  se puede considerar del tipo Markov y para la densidad de probabilidad condicional se cumple la ecuación (4.21). Los respectivos coeficientes de la ecuación se hallan integrando la ecuación diferencial del modelo (1) dentro de los límites  $(t, t + \Delta t)$ , con lo cual se obtiene,

$$\Delta x = m \int_t^{t+\Delta t} \eta(\lambda) d\lambda - \alpha \int_t^{t+\Delta t} \xi(\lambda) d\lambda$$

entonces,

$$E\{\Delta x\} = E\left\{\int_t^{t+\Delta t} \eta(\lambda) d\lambda\right\} - \alpha E\left\{\int_t^{t+\Delta t} \xi(\lambda) d\lambda\right\} = -\alpha \int_t^{t+\Delta t} E\{\xi(\lambda)\} d\lambda = -\alpha x \Delta t$$

De manera similar, se obtiene el segundo momento,

$$\begin{aligned} E\{\Delta t^2|x\} &= E\left\{m^2 \int_t^{t+\Delta t} \int_t^{t+\Delta t} \eta(\lambda_1) \eta(\lambda_2) d\lambda_1 d\lambda_2\right\} = m^2 \int_t^{t+\Delta t} \int_t^{t+\Delta t} E\{\eta(\lambda_1) \eta(\lambda_2)\} d\lambda_1 d\lambda_2 \\ &= m^2 \int_t^{t+\Delta t} \int_t^{t+\Delta t} R_\xi(\lambda_2 - \lambda_1) d\lambda_1 d\lambda_2 = m^2 \int_t^{t+\Delta t} \int_t^{t+\Delta t} \delta(\lambda_2 - \lambda_1) d\lambda_1 d\lambda_2 = m^2 \Delta t \end{aligned}$$

De las anteriores relaciones, teniendo en cuenta las respectivas definiciones (4.22), resultan los siguientes coeficientes  $a(x, t) = -\alpha x$  y  $b(x, t) = m^2$ .

**Ejemplo 4.9.** Hallar la solución de la ecuación diferencial, obtenida en el ejemplo 4.8, para la FDP condicional del proceso  $\xi(t)$  en el momento  $\tau$  con condiciones iniciales  $\xi(0) = x$ .

La solución de la ecuación 4.21, en general, depende de la forma de cambio de las funciones coeficiente (4.22). Una aproximación común supone la siguiente relación lineal de ambos coeficientes (pero asumiendo su estacionariedad):  $a(t, x) = \alpha_0 + \alpha_1 x$  y  $b(t, x) = \beta_0$ , donde  $\alpha_i, \beta_i, i = 1, 2$ , constantes. En este caso, es preferible el análisis realizarlo, no con la FDP, sino mediante la respectiva función característica, descrita en (2.17), con lo cual se obtiene la siguiente ecuación en derivadas parciales de primer orden:

$$\frac{\partial \Theta}{\partial t} + \left( \frac{1}{2} \beta_0 \omega^2 - j \alpha_0 \omega \right) - \alpha_1 \omega \frac{\partial}{\partial \omega} (\ln \Theta) = 0$$

que tiene solución,

$$\ln \Theta = -\frac{\sigma_\xi^2(t)}{2} \omega^2 + j \omega \overline{\xi(t)}$$

con coeficientes,

$$\begin{cases} \overline{\xi(t)} = \frac{\alpha_0}{\alpha_1} (e^{\alpha_1 t} - 1) + x e^{\alpha_1 t} \\ \sigma_\xi^2(t) = \frac{\beta_0}{2\alpha_1} (e^{2\alpha_1 t} - 1) \end{cases} \quad (2)$$

para las condiciones iniciales:  $\sigma_\xi(t) = 0, \overline{\xi(0)} = x$ .

De acuerdo con los resultados obtenidos en el ejemplo 4.8, entonces  $\alpha_0 = 0, \alpha_1 = -\alpha$  y  $\beta_0 = m^2$ , cuyos valores se reemplazan en la ecuación (2), luego, se obtiene que el proceso  $\xi(t)$  es de estructura Gaussiana con momentos,

$$\overline{\xi(t)} = x \exp(-\alpha t), \quad (3a)$$

$$\sigma_\xi^2(t) = \frac{m^2}{2\alpha} (1 - \exp(2\alpha t)) \quad (3b)$$

La varianza (3b) corresponde a la expresión de la varianza que se obtiene en el ejemplo 5.3 para el RBG a la salida del circuito  $RC$ , descrito por una ecuación diferencial con modelo similar a (1) en el ejemplo 4.8. En otras palabras, tanto el análisis de correlación, como el análisis de modelos de Markov convergen, sin embargo, en el último caso, se puede hacer una interpretación más profunda de los resultados, en particular, se puede afirmar que un proceso de Markov, con función de correlación  $R_\eta(t, \tau) = \frac{m^2}{2\alpha} e^{\alpha|\tau|}$ , también es de estructura Gaussiana. Viceversa, si un proceso Gaussiano estacionario tiene función de correlación  $R_\eta(t, \tau) = \frac{m^2}{2\alpha} e^{\alpha|\tau|}$ , entonces también es del tipo Markoviano.

**Descripción generalizada.** La descripción general de la dinámica estadística de un proceso continuo aleatorio  $x(t)$  con dimensión  $n$  en el tiempo se da por la ecuación diferencial:

$$\frac{d}{dt} x(t) = f(x, t) + \xi(x, t), \quad x(t_0) = x_0 \quad (4.23)$$

donde  $f(x, t)$  es el vector función no aleatorio con dimensión  $n$  y  $\xi(x, t)$  es el proceso aleatorio con propiedades probabilísticas conocidas, que en principio pueden depender del

vector  $\mathbf{x}$ . La expresión (4.23) se describe por medio de diferenciales en la forma:

$$d\mathbf{x}(t) = \mathbf{f}(\mathbf{x}, t)dt + d\boldsymbol{\eta}(\mathbf{x}, t), \quad \mathbf{x}(t_0) = \mathbf{x}_0 \quad (4.24)$$

donde  $d\boldsymbol{\eta}$  es el diferencial de algún proceso aleatorio  $d\boldsymbol{\eta}(\mathbf{x}, t)$  relacionado con el proceso aleatorio  $\boldsymbol{\xi}(\mathbf{x}, t)$ . Se puede demostrar [30], que siendo solución de la anterior ecuación diferencial, el proceso  $\mathbf{x}(t)$  es del tipo Markoviano, si al mismo tiempo  $\boldsymbol{\eta}(\mathbf{x}, t)$  es un proceso con incrementos (de cualquier tipo, positivos o negativos),  $\Delta\boldsymbol{\eta}(t_k) = \boldsymbol{\eta}(t_{k+1}) - \boldsymbol{\eta}(t_k)$ , que sean estadísticamente independientes. El proceso con incrementos independientes es del tipo homogéneo si se cumple que  $\Delta\boldsymbol{\eta}(t_k, t) = \Delta\boldsymbol{\eta}(t_k)$ , para los cuales la FDP es infinitamente divisible, lo que significa que si un intervalo de tiempo  $[t_0, t]$  se divide en  $N$  segmentos de igual duración  $\Delta t$ , entonces, para la función característica de los incrementos del proceso  $\boldsymbol{\eta}$  tiene lugar la igualdad

$$\Theta(j\boldsymbol{\vartheta}, t - t_0) = (\Theta(j\boldsymbol{\vartheta}, \Delta t))^N$$

En general, cualquier variable aleatoria, con FDP infinitamente divisible, se puede representar en forma de una suma de dos componentes independientes: una con estructura Gaussiana y la otra con estructura de Poisson,

$$\ln \Theta(j\boldsymbol{\vartheta}, \Delta t) = \Delta t \left( j\boldsymbol{\vartheta}^\top \mathbf{m}_1 - \frac{1}{2} \boldsymbol{\vartheta}^\top \mathbf{Q} \boldsymbol{\vartheta} + \int \left( e^{j\boldsymbol{\vartheta}^\top \mathbf{y}} - 1 \right) \pi(\mathbf{y}) d\mathbf{y} \right)$$

La primera componente,  $\Delta\dot{\mathbf{v}}$ , se puede analizar como la derivada de un proceso continuo Gaussiano, descrito tanto por el vector  $\mathbf{m}_1 \Delta t$  de valores medios con dimensión  $\mathbf{n}$  como por la matriz  $\mathbf{Q}_{n \times n} \Delta t$  de varianzas, que corresponde a un proceso  $\mathbf{v}(t)$  del tipo Wiener. Asumiendo  $\mathbf{m}_1 = 0$ , sin pérdida de generalidad, entonces, la componente Gaussiana se puede escribir en la forma:

$$\Delta\dot{\mathbf{v}}(t) = \mathbf{G} \Delta\mathbf{v}(t)$$

siendo  $\mathbf{G}_{n \times k}$ , tal que  $\mathbf{G} \mathbf{G}^\top = \mathbf{Q}$ , mientras  $\mathbf{v}(t)$  es el proceso Wiener con dimensión  $\mathbf{k}$ , de valor medio cero y matriz de varianza  $\mathbf{I} \Delta t$ .

La componente de Poisson,  $\Delta\boldsymbol{\gamma}(t)$ , que corresponde a los incrementos independientes del proceso, se puede representar como:

$$\Delta\boldsymbol{\gamma}(t) = \int \mathbf{C}(\mathbf{y}, t) \nu(d\mathbf{y}, \Delta t)$$

donde  $\mathbf{C}(\mathbf{y}, t)$  es una función determinada con dimensión  $\mathbf{n}$ , que corresponde a los valores de amplitud de los cambios o saltos que pueda tomar el proceso  $\Delta\boldsymbol{\gamma}(t)$ ;  $\nu(\mathbf{A}, \Delta t)$  es una variable aleatoria, que toma los valores  $0, 1, \dots, k$  con FDP Poisson:

$$P(\nu = k) = \frac{(\lambda(\mathbf{A}) \Delta t)^k}{k!} \exp(-\lambda(\mathbf{A}) \Delta t), \quad \lambda(\mathbf{A}) = \int_{\mathbf{A}} \pi(\mathbf{y}) d\mathbf{y}$$

Se puede demostrar que la componente Gaussiana del proceso con incrementos independientes es del tipo *estocástico continuo*, esto es, se cumple la relación,

$$\lim_{\Delta t \rightarrow 0} P(\|\eta(t + \Delta t) - \eta(t)\| > \varepsilon) = 0, \quad \forall \varepsilon > 0$$

Mientras, que para la componente de Poisson esta relación no se cumple (*Proceso estocástico con incrementos*).

A partir de lo anterior, se obtiene la descripción generalizada de la dinámica de los procesos continuos de Markov, por medio de las *ecuaciones estocásticas diferenciales*:

$$\frac{d}{dt} \mathbf{x}(t) = \mathbf{f}(\mathbf{x}, t) + \mathbf{G}(\mathbf{x}, t) \mathbf{N}(t) + \mathbf{r}(t) \quad (4.25a)$$

$$\begin{aligned} \dot{\mathbf{x}}(t) &= \mathbf{f}(\mathbf{x}, t) dt + \mathbf{G}(\mathbf{x}, t) d\mathbf{v}(t) + \dot{\mathbf{r}}(t) \\ &= \mathbf{f}(\mathbf{x}, t) dt + \mathbf{G}(\mathbf{x}, t) d\mathbf{v}(t) + \int \mathbf{C}(\mathbf{x}, \mathbf{y}, t) \nu(d\mathbf{y}, dt) \end{aligned} \quad (4.25b)$$

donde  $\nu(\mathbf{A}, \Delta t)$  se caracteriza por la función  $\pi(\mathbf{y} | t, \mathbf{x})$ , mientras,  $\mathbf{r}(t) = \dot{\mathbf{r}}(t)$  es el proceso, que corresponde la derivada en el tiempo de la componente de Poisson, y la cual se puede interpretar como una serie de pulsos  $\delta$ :

$$\mathbf{r}(t) = \sum_k \mathbf{C}(\mathbf{y}_k) \delta(t - t_k) \quad (4.26)$$

siendo  $\{t_k\}$ , el conjunto de valores aleatorios de los instantes de tiempo en los que ocurren los pulsos. La ecuaciones (4.25a) y (4.25b) se entienden en el *sentido de Stratanovich* [31], en las que aparecen el proceso aleatorio  $\mathbf{N}(t)$  con varianza infinita y un proceso  $\mathbf{r}(t)$ , el cual toma valores infinitos en los momentos de definición  $t = t_k$ , a diferencia de lo que ocurre en una ecuación diferencial ordinaria de la forma  $\dot{\mathbf{x}}(t) = \mathbf{f}(\mathbf{x}, t)$ , con condición inicial  $\mathbf{x}(t_0) = \mathbf{x}_0$ , y para la cual se supone la solución,

$$\mathbf{x}(t) = \mathbf{x}_0 + \int_{t_0}^t \mathbf{f}(\mathbf{x}, \tau) d\tau$$

entonces, la integral se entiende en el *sentido de Cauchy-Riemann* y se da por el límite de la suma, el cual existe si la función  $\mathbf{f}(\mathbf{x}, \tau)$  cumple las condiciones de Lipschitz [1].

$$\int_{t_0}^t \mathbf{f}(\mathbf{x}, \tau) d\tau = \lim_{\Delta t \rightarrow 0} \sum_{k=0}^{N-1} \mathbf{f}(\mathbf{x}(t_i), t_i) \Delta t, \quad \Delta t = t_{k+1} - t_k$$

De manera similar, se asume la solución del ecuación diferencial estocástica (4.25b), concretamente en la forma:

$$\mathbf{x}(t) = \mathbf{x}_0 + \int_{t_0}^t \mathbf{f}(\mathbf{x}, \tau) d\tau + \int_{t_0}^t \mathbf{G}(\mathbf{x}, \tau) d\mathbf{v}(\tau) + \int_{t_0}^t \int \mathbf{C}(\mathbf{x}, \mathbf{y}, \tau) \nu(d\mathbf{y}, d\tau) \quad (4.27)$$

para la cual existen, por lo menos, dos formas diferentes de definición de la integral de los procesos aleatorios. La primera integral se puede definir, en analogía con la forma Cauchy-Riemann, si se asume su convergencia en el sentido del valor cuadrático medio, esto es, l.i.m.  $\xi_n = \xi$ , cuando se cumple  $\lim_{n \rightarrow \infty} E \left\{ (\xi_n - \xi)(\xi_n - \xi)^T \right\} = 0$ .

La integral de la componente de Poisson, en correspondencia con (4.26), se escribe como:

$$\int_{t_0}^t \int \mathbf{C}(\mathbf{x}, \mathbf{y}, \tau) \nu(d\mathbf{y}, d\tau) = \int_{t_0}^t \sum_k \mathbf{C}(\mathbf{x}, \mathbf{y}, \tau) \delta(\tau - t_k) d\tau$$

En cuanto a la integral de la componente Gaussiana de (4.27), ésta se puede definir de diferente manera. La primera forma corresponde a la *integral estocástica en el sentido de Ito*, entendida como la suma del límite de las sumas convergentes en el valor cuadrático medio del tipo:

$$\hat{\mathbf{S}}(t) = \int_{t_0}^t \mathbf{G}(\mathbf{x}, \tau) d\hat{\nu}(\tau) = \text{l.i.m.}_{N \rightarrow \infty} \sum_{k=0}^{N-1} \mathbf{G}(\mathbf{x}(t_k, t_k)) (\nu(t_{k+1}) - \nu(t_k))$$

La ecuación diferencial, para la cual la correspondiente integral se entiende en el sentido de Ito, se denomina *ecuación estocástica de Ito*, descrita como:

$$d\mathbf{x}(t) = \mathbf{A}(\mathbf{x}, t)dt + \mathbf{G}(\mathbf{x}, t)d\hat{\nu}(t) + \gamma(t) \quad (4.28a)$$

$$\frac{d}{dt}\mathbf{x}(t) = \mathbf{A}(\mathbf{x}, t) + \mathbf{G}(\mathbf{x}, t)\hat{\mathbf{N}}(t) + \mathbf{r}(t) \quad (4.28b)$$

Si un proceso de Markov,  $\mathbf{x}(t)$ , con dimensión  $n$  cumple la condición (4.28a), entonces el proceso  $\mathbf{z}(t) = \Phi(\mathbf{x}, t)$ , siendo  $\Phi(\mathbf{x}, t)$  cierto vector funcional determinado con dimensión  $L$ , cumple la ecuación conocida como la *expresión de Ito*:

$$\begin{aligned} dz(t) = & \left\{ \frac{\partial}{\partial t} \Phi(\mathbf{x}, t) + \left( \frac{\partial}{\partial \mathbf{x}} \Phi(\mathbf{x}, t) \right) \mathbf{A}(\mathbf{x}, t) + \right. \\ & \left. \frac{1}{2} \left( \sum_m^N \sum_n^N \sum_l^K \frac{\partial^2 \Phi_i(\mathbf{x}, t)}{\partial x_m \partial x_n} G_{ml}(\mathbf{x}, t) G_{nl}(\mathbf{x}, t) \right)_{i=1, \dots, L} \right\} dt + \\ & + \left\{ \frac{\partial}{\partial \mathbf{x}} \Phi(\mathbf{x}, t) \right\} \mathbf{G}(\mathbf{x}, t) d\nu(t) + \\ & + \int (\Phi(\mathbf{x} + \mathbf{C}(\mathbf{x}, \mathbf{y}, t), t) - \Phi(\mathbf{x}, t)) \nu(d\mathbf{y}, dt) \end{aligned}$$

donde  $\Phi_i(\mathbf{x}, t)$ ,  $i = 1, \dots, L$  son las componentes del vector  $\Phi(\mathbf{x}, t)$ ,  $G_{ml}(\mathbf{x}, t)$  y  $G_{nl}(\mathbf{x}, t)$  son las componentes de la matriz  $\mathbf{G}(\mathbf{x}, t)$ .

La expresión de Ito se puede simplificar en la forma

$$dz(t) = \frac{\partial \Phi}{\partial t} dt + \frac{\partial \Phi}{\partial x} dx(t) + \frac{1}{2} \left( \sum_{m,n=1}^N \sum_{l=1}^K \frac{\partial^2 \Phi_i}{\partial x_m \partial x_n} G_{ml} G_{nl} \right)_{i=1,\dots,L} dt + \int \left( \Phi(x + C(x, y, t), t) - \Phi(x, t) - \frac{\partial \Phi}{\partial x} C(x, y, t) \right) \nu(dy, dt)$$

de la que se observa que si un proceso de Markov  $x(t)$  es continuo ( $C(x, y, t) = 0$ ), entonces el diferencial del proceso  $z(t)$ , con precisión hasta el término no aleatorio, es proporcional al diferencial del proceso  $x(t)$ . A propósito, el tercer término de la última ecuación, hace imposible el cambio de variables en forma convencional, integrando por partes, o mediante otras técnicas de integración. Además, la integral estocástica de Ito, determinada para el sentido directo del tiempo, no coincide con la de tiempo inverso, cuando en la definición de  $\hat{S}$  la función  $G(x, t)$  se toma en el extremo opuesto del intervalo elemental.

En [31] se define la integral estocástica en el sentido de Stratanovich, que es relativamente simétrica con respecto al pasado y futuro, y que corresponde al límite de las sumas convergentes en el valor cuadrático medio de la forma:

$$S(t) = \int_{t_0}^t G(x, \tau) d\nu(\tau) = \lim_{N \rightarrow \infty} \sum_{i=0}^{N-1} G\left(\frac{x(t_{i+1}) + x(t_i)}{2}, t_i\right) (\nu(t_{i+1}) - \nu(t_i))$$

En este caso, para la formación de las sumas integrales, los valores de  $G(X, \tau)$  se toman en los puntos medios de los subintervalos elementales. Una de las propiedades de la integral estocástica de Stratanovich está en la posibilidad de su empleo en las formas convencionales del análisis matemático. Las correspondientes ecuaciones (4.25a) y (4.25b), también se denominan *ecuaciones estocásticas en forma simétrica*. La integral simétrica se diferencia de la respectiva integral de Ito en el valor [31]:

$$S(t) - \hat{S}(t) = \frac{1}{2} \int_{t_0}^t \left( \sum_{m=1}^N \sum_{l=1}^K \frac{\partial G_{il}(x, \tau)}{\partial x_m} G_{ml}(x, \tau) \right) d\tau, \quad i = 1, \dots, N$$

esto es, exceptuando el caso cuando  $G(x, \tau)$  no depende de  $x$ , ambos integrales se diferencian, por esto es importante en la descripción de las ecuaciones estocásticas diferenciales, determinar en que forma se deben entender las ecuaciones. De otra parte, de la última ecuación resulta que las expresiones (4.25a) y (4.25b) en el sentido de Stratanovich y las respectivas (4.28a) y (4.28b) en el sentido de Ito, describen un mismo proceso de Markov, cuando se cumple la igualdad,

$$A_i(x, t) = f_i(x, t) + \frac{1}{2} \sum_{m=1}^N \sum_{l=1}^K \frac{\partial G_{il}(x, t)}{\partial x_m} G_{ml}(x, t) \quad (4.29)$$

donde  $A_i(x, t)$  y  $f_i(x, t)$  son las respectivas componentes de los vectores funciones.

La descripción de los procesos continuos de Markov incluye la definición de la FDP de transición  $\pi(\mathbf{x}, t \mid \mathbf{x}_0, t_0)$  y la FDP marginal  $p(\mathbf{x}, t)$ , relacionadas por la expresión:

$$p(\mathbf{x}, t) = \int_{\mathbf{x} \in X} \pi(\mathbf{x}, t \mid \mathbf{x}_0, t_0) p(\mathbf{x}_0, t_0) d\mathbf{x}_0 \quad (4.30)$$

que cumplen las condiciones  $\int_{\mathbf{x} \in X} p(\mathbf{x}, t) d\mathbf{x} = 1$ ,  $\int_{\mathbf{x} \in X} \pi(\mathbf{x}, t \mid \mathbf{x}_0, t_0) d\mathbf{x} = 1$ .

Además, para los procesos continuos de Markov es cierta la *ecuación de Kolmogorov-Chapman*:

$$\pi(\mathbf{x}, t \mid \mathbf{x}_0, t_0) = \int_{\mathbf{x} \in X} \pi(\mathbf{x}, t \mid \mathbf{x}_1, t_1) \pi(\mathbf{x}_1, t_1 \mid \mathbf{x}_0, t_0) d\mathbf{x}_1$$

donde  $X$  es el dominio de los valores del proceso  $\mathbf{x}(t)$ ,  $t > t_1, t_0$ .

Se puede demostrar [32], que si todos los elementos del vector  $\mathbf{A}(\mathbf{x}, t)$  son diferenciables, así mismo los elementos de  $\mathbf{G}(\mathbf{x}, t)$  son dos veces diferenciables por  $\mathbf{x}$ , entonces la FDP de transición  $\pi(\mathbf{x}, t \mid \mathbf{x}_0, t_0)$ , dada en función de  $\mathbf{x}$  y  $t$ , cumple la *ecuación directa de Kolmogorov-Feller*:

$$\frac{\partial}{\partial t} \pi(\mathbf{x}, t \mid \mathbf{x}_0, t_0) = \mathcal{K}_{t,x} \{ \pi(\mathbf{x}, t \mid \mathbf{x}_0, t_0) \} \quad (4.31)$$

con condiciones iniciales  $\pi(\mathbf{x}, t \mid \mathbf{x}_0, t_0) = \delta(\mathbf{x} - \mathbf{x}_0)$ . El operador  $\mathcal{K}_{t,x} \{ \cdot \}$ , denominado *primer operador derivante* se determina por la suma de los operadores

$$\mathcal{K}_{t,x} \{ \cdot \} = \mathcal{L}_{t,x} \{ \cdot \} + \mathcal{M}_{t,x} \{ \cdot \}$$

definidos por las respectivas expresiones:

$$\begin{aligned} \mathcal{L}_{t,x} \{ w(\mathbf{x}) \} &= - \sum_{i=1}^n \frac{\partial}{\partial x_i} (A_i(\mathbf{x}, t) w(\mathbf{x})) + \frac{1}{2} \sum_{i=1}^N \sum_{l=1}^N \frac{\partial^2}{\partial x_i \partial x_l} (B_{il}(\mathbf{x}, t) w(\mathbf{x})) \\ \mathcal{M}_{t,x} \{ w(\mathbf{x}) \} &= \int \int w(\xi) (\delta(\mathbf{x} - \xi - \mathbf{C}(\xi, \mathbf{y}, t)) - \delta(\mathbf{x} - \xi)) \pi(\mathbf{y} \mid t, \xi) d\mathbf{y} d\xi \end{aligned}$$

donde  $A_i(\mathbf{x}, t)$  se denominan la *deriva del proceso* y se determinan de la expresión (4.29), mientras  $B_{il}(\mathbf{x}, t)$  son elementos positivos que determinan la matriz denominada *matriz de difusión* de orden  $n \times n$  e igual a

$$\mathbf{B}(\mathbf{x}, t) = \mathbf{G}(\mathbf{x}, t) \mathbf{G}^T(\mathbf{x}, t)$$

La FDP de transición, como función de  $\mathbf{x}_0$  y  $t_0$ , cumple con la *ecuación inversa de Kolmogorov-Feller*,

$$-\frac{\partial}{\partial t} \pi(\mathbf{x}, t \mid \mathbf{x}_0, t_0) = \mathcal{K}_{t,x} \{ \pi(\mathbf{x}, t \mid \mathbf{x}_0, t_0) \} \quad (4.32)$$



mediante el operador inverso  $\mathcal{K}_{t,x}^{-1}\{\cdot\} = \mathcal{K}_{t,x}^*\{\cdot\}$ , que resulta ser su propio conjugado:

$$\begin{aligned}\mathcal{L}_{t,x}^*\{w(x)\} &= \sum_{i=1}^n A_i(x, t) \frac{\partial w(x)}{\partial x_i} + \frac{1}{2} \sum_{i=1}^N \sum_{l=1}^N B_{il}(x, t) \frac{\partial^2 w(x)}{\partial x_i \partial x_l} \\ \mathcal{M}_{t,x}^*\{w(x)\} &= \int (w(x + \mathbf{C}(x, \mathbf{y}, t)) - w(x)) \pi(\mathbf{y} | t, x) d\mathbf{y}\end{aligned}$$

Debido a la relación entre los operadores directo e inverso, se tiene que para cualquier dominio  $\Omega$ , en el cual los elementos de la función  $\mathbf{A}(x, t)$  son diferenciables, además los elementos de  $\mathbf{B}(x, t)$  son dos veces diferenciables por  $x$ , es cierta la relación:

$$\int_{\Omega} u(x) \mathcal{K}_{t,x}\{w(x)\} dx = \int_{\Omega} w(x) \mathcal{K}_{t,x}^*\{u(x)\} dx$$

donde  $u(x)$  y  $w(x)$  son funciones escalares, de las cuales por lo menos una con sus primeras derivadas son iguales a cero en la frontera del dominio  $\Omega$ .

La FDP marginal dada en (4.30), teniendo en cuenta (4.31), corresponde a la ecuación

$$\frac{\partial}{\partial t} p(x, t) = \mathcal{K}_{t,x}\{p(x, t)\} \quad (4.33)$$

que se debe resolver para la condición inicial  $p(x, t) = p(x_0)$ .

Las ecuaciones (4.31), (4.32) y (4.33) son del tipo integro-diferenciales con derivadas parciales, para las cuales no hay una solución compacta e inclusive para el caso estacionario la solución es muy compleja.

Si no existe la componente de Poisson en un proceso con incrementos independientes en la ecuaciones diferenciales estocásticas (4.25b) y (4.28a), entonces el proceso continuo de Markov  $x(t)$  se denomina *difusivo*, para el cual las ecuaciones (4.31), (4.32) y (4.33) se denominan ecuaciones de *Fokker-Plank-Kolmogorov*. En los procesos difusivos de Markov se introduce el concepto de *flujo de la densidad de probabilidad*  $\Pi(x, t)$ , cuyas componentes en cualquier punto de  $x$  son iguales a

$$\Pi_i(x, t) = A_i(x, t) + p(x, t) - \frac{1}{2} \sum_{l=1}^N \frac{\partial}{\partial x_l} (B_{il}(x, t) p(x, t)), \quad i = 1, \dots, N \quad (4.34)$$

La ecuación directa de Fokker-Plank-Kolmogorov (4.34) usualmente se escribe como:

$$\frac{\partial}{\partial t} p(x, t) = \mathcal{L}_{t,x}\{p(x, t)\} = - \sum_{i=1}^N \frac{\partial}{\partial x_i} \Pi_i(x, t) = -\text{div} \Pi(x, t)$$

la cual se interpreta como la ley de conservación de la función densidad de probabilidad.

En forma general, se consideran diferentes tipos de comportamiento de un proceso difusivo de Markov, con respecto a las partículas que se difunden en la frontera del dominio, entre otros, los siguientes: absorción, reflexión, salida abrupta de la frontera, parada, etc.

El cálculo numérico de las ecuaciones de Fokker-Plank-Kolmogorov se realiza empleando

diferentes métodos de aproximación, entre ellos, los de Galerkin, iterativos, esperanzas matemáticas condicionales, y en particular, el método de Monte-Carlo, que se basa en la solución numérica de las ecuaciones diferenciales estocásticas (4.25a) y (4.25b) [30].

#### 4.1.4. Sucesiones de Markov

Corresponde a sucesiones vectores de variables aleatorias,  $\mathbf{x}_i = \mathbf{x}(t_i)$ , con dimensión  $n$ , las cuales, en determinados momentos del tiempo,  $t_0 < t_1 < \dots < t_i < \dots$ , forman una escala continua de posibles valores, que cumplen las condiciones (4.1) y (4.2). La discretización de procesos continuos aleatorios con carga informativa es un ejemplo en el cual, empleando las sucesiones de Markov, se puede analizar la dinámica estadística de los diferentes sistemas de procesamiento digital de procesos aleatorios.

De la expresión (4.2), se observa que la FDP conjunta de una sucesión de Markov puede ser expresada a través de la densidad de probabilidad del estado inicial  $p(\mathbf{x}_0)$  y la densidades de probabilidad de cambio elemental  $\pi_i(\mathbf{x}_i | \mathbf{x}_{i-1})$ ,

$$p(\mathbf{x}_0, \dots, \mathbf{x}_k) = p(\mathbf{x}_0) \prod_{i=1}^k \pi_i(\mathbf{x}_i | \mathbf{x}_{i-1})$$

En analogía con las cadenas compuestas de Markov, se pueden definir las sucesiones compuesta de Markov de orden  $m > 1$ , de tal manera que se cumple

$$\pi_k(\mathbf{x}_k | \mathbf{x}_0, \dots, \mathbf{x}_{k-1}) = \pi_k(\mathbf{x}_k | \mathbf{x}_{k-m}, \dots, \mathbf{x}_{k-1})$$

donde  $k \leq m$ . También en este caso cualquier sucesión compuesta se puede representar mediante una combinación de cadenas simples.

Cualquier sucesión, escogida dentro de otra sucesión de Markov, vuelve a ser del tipo Markov, esto es, si para un momento dado  $t_k$  se analiza la malla de valores discretos tiempos  $t_{k_1} < t_{k_2} < \dots < t_{k_m}$ , entonces  $\pi(\mathbf{x}_{k_m} | \mathbf{x}_{k_1}, \dots, \mathbf{x}_{k_{m-1}}) = \pi_{k_m}(\mathbf{x}_{k_m} | \mathbf{x}_{k_{m-1}})$ .

Las densidades condicionales de probabilidad de transición para una sucesión de Markov cumplen con la ecuación de Kolmogorov-Chapman

$$\pi(\mathbf{x}_l | \mathbf{x}_i) = \int_{\mathbf{x} \in X} \pi(\mathbf{x}_l | \mathbf{x}_k) \pi(\mathbf{x}_k | \mathbf{x}_i) d\mathbf{x}_k \quad (4.35)$$

donde  $i < k < l$  y  $X$  es el dominio de los valores del proceso  $\mathbf{x}(t)$ . Si se nota por  $p_k(\mathbf{x} | \mathbf{x}_0)$  la densidad condicional de la probabilidad de la variable aleatoria  $\mathbf{x}_k$ , para un valor fijo de  $\mathbf{x}_0$  y considerando  $\pi_k(\mathbf{x} | \mathbf{z}) = \pi_k(\mathbf{x}_k | \mathbf{x}_{k-1})$ , de (4.35) se obtiene que

$$p_{k+1}(\mathbf{x} | \mathbf{x}_0) = \int_X \pi_k(\mathbf{x} | \mathbf{z}) p_k(\mathbf{z} | \mathbf{x}_0) d\mathbf{z} \quad (4.36)$$

Así, si se conocen las FDP elementales de transición  $\pi_k(\mathbf{x} | \mathbf{z})$ , entonces la relación recurrente (4.36) permite calcular las características estadísticas de una sucesión de Markov.

Una sucesión de Markov se considera *homogénea* si las FDP elementales de transición  $\pi_k(\mathbf{x} \mid \mathbf{z})$  no dependen de  $k$ , así mismo, se considera *estacionaria* si la cadena, además de ser homogénea, todos los estados  $\mathbf{x}_k$  tienen una sola FDP,  $p(\mathbf{x}) = \lim_{k \rightarrow \infty} p_k(\mathbf{x})$ , que de existir, cumple la ecuación integral:

$$p(\mathbf{x}) = \int_X \pi(\mathbf{x} \mid \mathbf{z}) p(\mathbf{z}) d\mathbf{z}, \quad \pi(\mathbf{x} \mid \mathbf{z}) = \lim_{k \rightarrow \infty} \pi_k(\mathbf{x} \mid \mathbf{z}) \quad (4.37)$$

Similar a (4.23) y (4.24), la descripción generalizada de la dinámica estadística de una sucesión de Markov  $\mathbf{x}_k = \mathbf{x}(t_k)$  con dimensión  $n$ , en tiempo discreto, se da como,

$$\mathbf{x}_{k+1} = \Phi_k(\mathbf{x}_k, \mathbf{w}_k) \quad (4.38)$$

donde  $\Phi_k(\mathbf{x}_k, \mathbf{w}_k)$  es la función vectorial que no es aleatoria con respecto a sus argumentos;  $\mathbf{w}_k = \mathbf{w}(t_k)$  son los valores independientes del proceso aleatorio discretizado con FDP conocidas  $q_k(\mathbf{w}) = q(\mathbf{w}, t_k)$ . Cabe anotar que la ecuación (4.38) determina la relación entre las variables aleatorias  $\mathbf{x}_{k+1}$  y  $\mathbf{x}_k$  para valores fijos de  $\mathbf{x}_k$ . Por esta razón, las FDP elementales de transición que participan en las ecuaciones (4.36) y de (4.37), se hallan de (4.38) empleando diferentes reglas de transformación [15].

En el análisis de las sucesiones de Markov, también se pueden dar diferentes comportamientos en las fronteras de un dominio dado  $\omega \subset X$ . Por ejemplo, si es de interés hallar la primera salida de una sucesión homogénea de Markov fuera de la frontera  $\Gamma \subset \Omega$ , entonces la probabilidad  $P(k, \mathbf{x}_0)$  de que la sucesión de Markov, comenzando de  $\mathbf{x}_0 \in \Omega$ , salga por primera vez de la frontera  $\Gamma$  en el estado de transición  $k$  se halla de la relación:

$$P(k, \mathbf{x}_0) = \int_{\Omega} (p_{k-1}(\mathbf{x} \mid \mathbf{x}_0) - p_k(\mathbf{x} \mid \mathbf{x}_0)) d\mathbf{x} \quad (4.39)$$

donde la densidad de probabilidad  $p_k(\mathbf{x} \mid \mathbf{x}_0)$  cumple la ecuación recurrente

$$p_k(\mathbf{x} \mid \mathbf{x}_0) = \int_{\Omega} \pi(\mathbf{x} \mid \mathbf{z}) p_{k-1}(\mathbf{z} \mid \mathbf{x}_0) d\mathbf{z} \quad (4.40)$$

De (4.39), la cantidad media de transiciones hasta la primera salida fuera de la frontera  $\Gamma \subset \Omega$  es

$$E\{k \mid \mathbf{x}_0\} = \sum_{k=1}^{\infty} k P(k, \mathbf{x}_0) = 1 + \int_{\Omega} \sum_{k=1}^{\infty} p_k(\mathbf{x} \mid \mathbf{x}_0) d\mathbf{x} = 1 + \int_{\Omega} P(\mathbf{x} \mid \mathbf{x}_0) d\mathbf{x}$$

siendo  $P(\mathbf{x} \mid \mathbf{x}_0) = \sum_{k=1}^{\infty} p_k(\mathbf{x} \mid \mathbf{x}_0)$ . la misma función  $P(\mathbf{x} \mid \mathbf{x}_0)$ , como se observa de (4.40) se puede hallar de la solución de la ecuación integral

$$P(\mathbf{x} \mid \mathbf{x}_0) = \pi(\mathbf{x} \mid \mathbf{x}_0) + \int_{\Omega} (\mathbf{x} \mid \mathbf{z}) P(\mathbf{z} \mid \mathbf{x}_0) d\mathbf{z}$$

El cálculo de la última ecuación integral se puede realizar empleando los métodos conocidos de solución para ecuaciones integrales homogéneas y no homogéneas de Fredholm.

## Problemas

**Problema 4.1.** Dadas las matrices de transición

$$\Pi = \begin{bmatrix} 0 & 2/3 & 1/3 \\ 1/2 & 0 & 1/2 \\ 1/3 & 2/3 & 0 \end{bmatrix}; \quad \Pi = \begin{bmatrix} 1/2 & 1/3 & 1/6 \\ 1/2 & 1/3 & 1/6 \\ 1/2 & 1/3 & 1/6 \end{bmatrix}, \quad \Pi = \begin{bmatrix} 0 & 1/2 & 1/2 \\ 1/2 & 0 & 1/2 \\ 1/2 & 1/2 & 0 \end{bmatrix}$$

Calcular las respectivas matrices de transición para  $n = 2, 3, 4, 5$  pasos.

**Problema 4.2.** Sea  $\xi_t$  el número de orden de un estado en la cadena de Markov en el instante de tiempo  $t$ ,  $P(\xi_0 = 1) = 1$ , y la matriz de probabilidades de transición tiene la forma

$$\Pi = \begin{bmatrix} 3/7 & 3/7 & 1/7 \\ 1/11 & 2/11 & 8/11 \\ 1/11 & 4/11 & 6/11 \end{bmatrix}$$

Demostrar que la sucesión derivada  $\eta_t = \begin{cases} 1, & \xi_t = 1 \\ 2, & \xi_t \neq 1 \end{cases}$  es un cadena de Markov y hallar su respectiva matriz de transición.

**Problema 4.3.** Un electrón puede encontrarse en una de tres órbitas con una probabilidad de cambio en unidad de tiempo de la órbita  $i$  a la  $j$  igual a,  $C_i \exp(-\lambda|i-j|)$ ,  $\lambda > 0$ . Hallar las probabilidades de cambio en dos unidades de tiempo y determinar la constante  $C_i$ .

**Problema 4.4.** Dada la matriz de transición

$$\begin{bmatrix} 0 & 1/2 & 1/2 \\ 1/4 & 1/4 & 1/4 \\ 1/2 & 1/2 & 0 \end{bmatrix}$$

establecer si es ergódica o no y calcular sus probabilidades finales.

**Problema 4.5.** Dada la matriz

$$\Pi = \begin{bmatrix} 1 & 0 \\ 1/3 & 2/3 \end{bmatrix}$$

Demostrar que la matriz  $\Pi^n$ , para  $n \rightarrow \infty$  no cumple el principio de ergodicidad, sin embargo, la matriz se puede considerar ergódica. Explicar porqué y hallar las probabilidades finales.

**Ejercicio en el CP 4.1.** Realizar un programa de realice, en el caso de las cadenas de Markov, los siguientes procedimientos:

1. Cálculo de las expresiones (4.7), dados los respectivos vectores y matrices iniciales.
2. Resolución del sistema algebraico de ecuaciones lineales (4.9).
3. Análisis de la convergencia de los elementos de transición en los valores de probabilidad final para el caso de la sucesiones estacionarias de Markov, por ejemplo en forma de un espectrograma.

**Problema 4.6.** Hallar la probabilidad de estado  $P_k(t)$  para un proceso homogéneo de Markov, en el cual se tiene  $\alpha_k = \alpha$  y  $\beta_k = k\beta$ ,  $k = 0, \dots, K$ . Resolver la ecuación diferencial asumiendo la ergodicidad del proceso aleatorio.

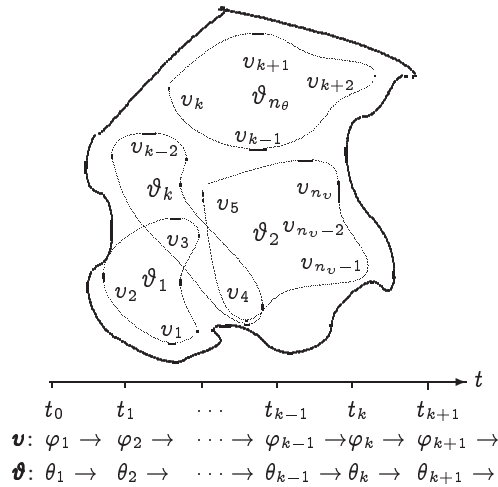
## 4.2. Análisis experimental de procesos de Markov

En el proceso real de señales, independiente de su forma de representación, análoga o discreta, no siempre está dado su modelo de aleatoriedad, y si está dado se debe resolver el problema del cálculo de sus características. La importancia de la caracterización de los modelos de las señales está en su papel básico durante la implementación de sistemas de proceso, entre ellos, los de reconocimiento, identificación, predicción, etc.

Los modelos de aleatoriedad de Markov, que se describen en el numeral §4.1, son de naturaleza probabilístico y su empleo adecuado en tareas de aplicación implica la estimación y ajuste de sus parámetros, que se realiza mediante el análisis estadístico de las señales reales. Cabe anotar, que debido a la complejidad en el análisis de los modelos de Markov, la caracterización de las señales aleatorias es preferible realizarla mediante métodos de estimación paramétrica.

### 4.2.1. Modelos ocultos de Markov

En general, las sucesiones de Markov, descritas en el numeral §4.1.4, asumen una cantidad finita de valores discretos o estados para la representación de la señal aleatoria. En particular, cada estado de manera directa se asocia a un evento físico observable. Sin embargo, en la práctica, se tienen aplicaciones con señales que no presentan de forma evidente los eventos sobre los cuales se construye el modelo. En este sentido, se debe construir un modelo probabilístico sobre los estados no observables u *ocultos*. Como resultado las cadenas construidas por este principio, corresponden a un proceso estocástico doblemente inmerso; la función probabilística de los estados ocultos y el mismo modelo de aleatoriedad de Markov impuesto sobre el proceso.



**Figura 4.4.** Representación de estados en los modelos ocultos de Markov

Los modelos ocultos de Markov, cuyos posibles estados se representan en la Figura 4.4, se pueden caracterizar mediante los siguientes parámetros:

- (a). El número de símbolos de observación o volumen  $n_v$  del alfabeto discreto de observación  $\mathbf{v} = \{v_k : k = 1, \dots, n_v\} \in \mathcal{U}$ . Los símbolos de observación corresponden a la salida física del sistema en análisis y conforman la sucesión aleatoria  $\boldsymbol{\varphi} = \{\varphi_1, \dots, \varphi_{n_\varphi}\}$  en los momentos definidos de tiempo  $1, 2, \dots, n_\varphi$ , donde  $n_\varphi$  es la longitud de la sucesión de observación.
- (b). El número de los estados ocultos del modelo,  $\boldsymbol{\vartheta} = \{\vartheta_k : k = 1, \dots, n_\vartheta\} \in \mathcal{V}$ , que siendo no observables, pueden ser relacionados con algún sentido físico del proceso. Los estados ocultos conforman la sucesión aleatoria  $\boldsymbol{\theta} = \{\theta_0, \theta_1, \dots, \theta_{n_\theta}\}$  en los momentos definidos de tiempo  $= 1, 2, \dots, n_\theta$ , donde  $n_\theta$  es la longitud de la sucesión de estados en análisis.

En general, los estados ocultos del modelo de cadenas de Markov se obtienen a partir del alfabeto de los símbolos de observación, por la transformación,  $\mathcal{V} = \mathcal{K} \{\mathcal{U}\}$ , tal que se conserven las medidas estadísticas:  $\sum_k^{n_\vartheta} P(\vartheta_k) = \sum_k^{n_v} P(v_k) = 1$ .

- (c). La matriz probabilidad de transición de estados,  $\boldsymbol{\Pi} = \{\pi_{mn} : m, n = 1, \dots, n_\vartheta\}$ , en la cual cada elemento se determina como,

$$\pi_{mn}(k) = P(\theta_{k+1} = \vartheta_n | \theta_k = \vartheta_m), \quad \pi_{mn} \geq 0, \quad \sum_{n=1}^{n_\vartheta} \pi_{mn} = 1$$

- (d). La matriz probabilidad condicional  $\mathbf{P}(\mathbf{v}|\boldsymbol{\vartheta})$  de los símbolos de observación respecto a cada estado  $\vartheta_i$  en el momento de tiempo  $n$ , con elementos,

$$P_{ki}(n) = P(v_k | \theta_n = \vartheta_i), \quad 1 \leq i \leq n_\vartheta, \quad 1 \leq k \leq n_v$$

- (e). El vector probabilidad de estado inicial  $\mathbf{p}_{\theta_1}$  con elementos  $\{P_{\theta_1}(i)\}$ , donde

$$p_{\theta_1}(i) = P(\theta_1 = \vartheta_i), \quad 1 \leq i \leq n_\vartheta$$

Los valores de aleatoriedad  $\boldsymbol{\Pi}$ ,  $\mathbf{P}(\mathbf{v}|\boldsymbol{\vartheta})$  y  $\mathbf{p}_{\theta_1}$ , notados en conjunto como

$$\boldsymbol{\mu} = \{\boldsymbol{\Pi}, \mathbf{P}(\mathbf{v}|\boldsymbol{\vartheta}), \mathbf{p}_{\theta_1}\}$$

conforman los parámetros de un modelo oculto de Markov, el cual se puede emplear para generar la estimación de la sucesión de observación,  $\boldsymbol{\varphi} \in \{\varphi_1, \varphi_2, \dots, \varphi_{n_\varphi}\}$ , con longitud  $n_\varphi = n_\vartheta$  para los momentos definidos de tiempo  $n = 1, 2, \dots, n_\varphi$ .

El desarrollo de las modelos ocultos de Markov está relacionado con las siguientes tres tareas estadísticas [33]:

1. Dada una sucesión de observación  $\boldsymbol{\varphi} = \{\varphi_1, \varphi_2, \dots, \varphi_{n_\varphi}\}$  con longitud  $n_\varphi$  y el modelo establecido,  $\boldsymbol{\mu} = \{\boldsymbol{\Pi}, \mathbf{P}(\mathbf{v}|\boldsymbol{\vartheta}), \mathbf{p}_{\theta_1}\}$ , cómo calcular de manera eficiente la probabilidad  $P(\boldsymbol{\varphi}|\boldsymbol{\mu})$  de la sucesión de observación.

2. Dada una sucesión de observación  $\boldsymbol{\varphi} = \{\varphi_1, \varphi_2, \dots, \varphi_{n_\varphi}\}$  con longitud  $n_\varphi$ , además, conocido el modelo  $\mu$ , cómo estimar de forma óptima, para una criterio de medida fijado a priori, la correspondiente sucesión de estados  $\boldsymbol{\theta} = \{\theta_1, \theta_2, \dots, \theta_{n_\varphi}\}$ .
3. El ajuste de los parámetros del modelo  $\mu$  que brinden el máximo valor de  $P(\boldsymbol{\varphi}|\mu)$ .

#### 4.2.2. Entrenamiento del modelo

**Estimación de la probabilidad  $P(\boldsymbol{\varphi}|\mu)$ .** Un primer método directo de cálculo de la probabilidad, para una sucesión de observación dada, consiste en la enumeración de cada una de los posibles sucesiones de estados que explican esa sucesión de observación.

Sea la sucesión  $i$  de estados ocultos con longitud  $n_\varphi$ ,  $\boldsymbol{\theta}(i) = \{\theta_1, \theta_2, \dots, \theta_n, \dots, \theta_{n_\varphi}\}$ , para la cual la probabilidad de la sucesión de observación, asumiendo la independencia de las observaciones, así como la independencia en el tiempo de las probabilidades  $\pi_{mn}$  y  $P_{ki}$ , está dada como

$$P(\boldsymbol{\varphi}|\boldsymbol{\theta}(i), \mu) = \prod_{n=1}^{n_\varphi} P(\varphi_n|\theta_n, \mu)$$

La probabilidad de que la sucesión dada de estados  $\boldsymbol{\theta}(i)$  ocurra se determina como,

$$P(\boldsymbol{\theta}(i)|\mu) = p_{\theta_1}(\theta_1) \pi_{\theta_1\theta_2} \cdots \pi_{\theta_{n_\varphi-1}\theta_{n_\varphi}}$$

de tal manera, que la probabilidad conjunta de que se tengan en forma simultanea ambas sucesiones corresponde al producto de los anteriores dos términos,

$$P(\boldsymbol{\varphi}, \boldsymbol{\theta}(i)|\mu) = P(\boldsymbol{\varphi}|\boldsymbol{\theta}(i), \mu)P(\boldsymbol{\theta}(i)|\mu) \quad (4.41)$$

La interpretación de la expresión (4.41) corresponde a que en el momento inicial de tiempo,  $n = 1$ , se tiene el estado  $\theta_1$  con una probabilidad de  $p_{\theta_1}(\theta_1)$  y se genera el símbolo de observación  $\varphi_1$  con una probabilidad de  $P_{\theta_1\varphi_1}$ . En el siguiente momento del tiempo, cuando  $n = 2$ , se realiza la transición al estado  $\theta_2$  desde  $\theta_1$  con una probabilidad de  $\pi_{\theta_1\theta_2}$  y se genera el símbolo  $\varphi_2$  con una probabilidad de  $P_{\theta_2\varphi_2}$ , y así sucesivamente, hasta el momento de tiempo  $n = n_\varphi$ . Finalmente, dado el modelo de aleatoriedad  $\mu$ , la probabilidad de que ocurra la sucesión de observación  $\boldsymbol{\varphi}$  se obtiene sumando la probabilidad conjunta (4.41) sobre todos las posibles sucesiones de estado  $\boldsymbol{\theta}(i)$ ,  $\forall i$ ,

$$P(\boldsymbol{\varphi}|\mu) = \sum_i P(\boldsymbol{\varphi}|\boldsymbol{\theta}(i), \mu)P(\boldsymbol{\theta}(i)|\mu) \quad (4.42)$$

La cantidad de operaciones en (4.42) es igual a  $(2n_\varphi - 1)n_\varphi^{n_\varphi}$  multiplicaciones y  $n_\varphi^{n_\varphi}$  adiciones, que hace el cálculo de la probabilidad  $P(\boldsymbol{\varphi}|\mu)$  prácticamente irrealizable. En este sentido, se emplean diferentes algoritmos recursivos que disminuyan el coste computacional.

El primer procedimiento corresponde al *algoritmo recursivo de propagación*, cuando se analiza el desarrollo de la cadena de estados en el sentido natural de desarrollo del tiempo,

y en el cual se realiza el desdoblamiento de las sucesiones de observación, empleando la relación inherente entre los elementos contiguos de las cadenas. En particular, para cada uno de los estados ocultos se puede definir la siguiente probabilidad parcial de una sucesión de observación,  $\{\varphi_1, \varphi_2, \dots, \varphi_n\}$  y del estado  $\vartheta_i$  en el momento de tiempo  $n = 1, 2, \dots$ , dado el modelo  $\mu$ ,  $\alpha_n(i) = P(\varphi_1, \varphi_2, \dots, \varphi_n, \theta_n = \vartheta_i | \mu)$ , con condición inicial para  $n = 1$ ,

$$\alpha_1(i) = p_{\theta_1}(i) P_{i\varphi_1}, \quad 1 \leq i \leq n_\vartheta$$

De tal manera, que para cada estado oculto tiene lugar la siguiente expresión recursiva

$$\alpha_{n+1}(j) = \left( \sum_{i=1}^{n_\vartheta} \alpha_n(i) \pi_{ij} \right) P_{j\varphi_{n+1}}, \quad n = 1, \dots, n_\varphi - 1, \quad 1 \leq j \leq n_\vartheta \quad (4.43)$$

La probabilidad  $P(\varphi | \mu)$  de generación de la sucesión de observación se obtiene sumando todas las probabilidades parciales con  $n = n_\varphi$  para todo el alfabeto de estados ocultos,

$$P(\varphi | \mu) = \sum_{i=1}^{n_\vartheta} \alpha_{n_\varphi}(i) \quad (4.44)$$

El algoritmo recursivo (4.43) se desarrolla en el sentido natural del tiempo y, por lo tanto, se denomina de *propagación directa*. Sin embargo, existe la versión equivalente de *propagación inversa* en el tiempo. Para esta propagación se define la probabilidad de la sucesión parcial de observación desde  $n + 1$  hasta  $n_\varphi$ , dado el estado  $\vartheta_i$  en el tiempo  $n$  y el modelo  $\mu$ ,  $\beta_n(i) = P(\varphi_{n+1}, \varphi_{n+2}, \dots, \varphi_{n_\varphi} | \theta_n = \vartheta_i, \mu)$ , con la debida condición final,  $\beta_{n_\varphi}(i) = 1$ ,  $1 \leq i \leq n_\vartheta$ .

Para cada estado oculto se tiene la siguiente expresión recursiva,

$$\beta_n(i) = \sum_{j=1}^{n_\vartheta} \pi_{ij} P_{j\varphi_{n+1}} \beta_{n+1}(j), \quad n = n_\varphi - 1, \dots, 1, \quad 1 \leq j \leq n_\vartheta \quad (4.45)$$

En cualquier caso, el algoritmo recursivo de propagación (directa o inversa) requiere tan solo de  $n_\vartheta^2 n_\varphi$  operaciones, que corresponde a un volumen aceptable de operaciones en condiciones reales de proceso.

**Selección de la mejor sucesión.** En este caso, no se tiene una solución que de forma universal, se acepte para la selección de la mejor sucesión de estados. La dificultad está en la definición de la misma sucesión óptima de estados.

Una primera solución consiste en la selección de los estados  $q_e$  que sean individualmente más probables; caso en el que se maximiza el número esperado de estados individuales correctos, para lo cual se define la variable,

$$\gamma_n(i) = P(\theta_n = \vartheta_i | \varphi, \mu) \quad (4.46)$$

esto es, la probabilidad de encontrarse en el estado  $\vartheta_i$  en el tiempo  $n$ , dada la observación



$\varphi$  y el modelo  $\mu$ . La ecuación (4.46) se expresa simplemente en términos de las variables de propagación:

$$\gamma_n(j) = \frac{\alpha_n(j)\beta_n(j)}{P(\varphi|\mu)} = \frac{\alpha_n(j)\beta_n(j)}{\sum_{i=1}^{n_\vartheta} \alpha_n(i)\beta_n(i)}$$

donde  $\alpha_n(j)$  describe la sucesión parcial de observación  $\varphi_1, \dots, \varphi_n$  y el estado  $\vartheta_j$  en tiempo  $n$ , mientras  $\beta_n(j)$  describe el resto de la sucesión de observación dado el estado  $\vartheta_j$  en  $n$ . El factor de normalización  $P(\varphi|\mu) = \sum_{i=1}^{n_\vartheta} \alpha_n(i)\beta_n(i)$ , le da el sentido de medida de probabilidad a  $\gamma_n(j)$ , de tal manera que,  $\sum_{i=1}^{n_\vartheta} \gamma_n(i) = 1$ .

Mediante el término  $\gamma_n(j)$  se puede hallar individualmente el estado  $\theta_n$  en el tiempo  $n$  más probable, como,

$$\theta_n = \arg \max_{1 \leq i \leq n_\vartheta} \{\gamma_n(i)\}, \quad 1 \leq n \leq n_\varphi \quad (4.47)$$

Aunque la expresión (4.47) maximiza el número esperado de estados correctos, (escogiendo el estado más probable por cada  $n$ ), podrían generarse algunos problemas con la sucesión de estados resultante. Por ejemplo, cuando el modelo oculto contempla transiciones de estado con valor de probabilidad cero, ( $\pi_{ij} = 0$ ), la sucesión de estados óptima ni siquiera podría ser una sucesión de estados válida. Este problema se debe a que la solución planteada en (4.47) simplemente determina el estado más probable en cada instante del tiempo, sin importar la probabilidad de ocurrencia de las sucesiones de estado.

Una posible corrección a este problema está en modificar el criterio de optimización, por ejemplo, se puede hallar la sucesión de estados que maximice el número esperado de pares de estados correctos  $(\theta_n, \theta_{n+1})$  o inclusive de cadenas más largas  $(\theta_n, \theta_{n+1}, \dots, \theta_{n+k})$ . Sin embargo, este criterio no ha mostrado ser implementable para un rango amplio de aplicaciones. Por lo tanto, en la práctica, el criterio con mayor uso se limita a hallar la mejor sucesión de estados, esto es, se maximiza  $P(\theta|\varphi, \mu)$ , considerando que es equivalente a maximizar  $P(\theta, \varphi|\mu)$ . La técnica formal de cálculo de la mejor sucesión de estados, basada en métodos de programación dinámica, corresponde al *algoritmo de Viterbi*.

Sea determinada la mejor sucesión de estados  $\theta = \{\theta_1, \dots, \theta_{n_\varphi}\}$ , para una sucesión de observación  $\varphi = \{\varphi_1, \dots, \varphi_{n_\varphi}\}$ . Se define la medida,

$$m_n(i) = \max_{\theta_1, \theta_2, \dots, \theta_{n-1}} P(\theta_1, \theta_2, \dots, \theta_n = i, \varphi_1, \varphi_2, \dots, \varphi_n | \mu)$$

que es la más alta probabilidad a lo largo de una única ruta, en el tiempo  $n$ , que da cuenta de las primeras  $n$  observaciones y finaliza en el estado  $\vartheta_i$ . De forma análoga se tiene:

$$m_{n+1}(j) = \left( \max_i m_n(i) \pi_{ij} \right) P_{j\varphi_{n+1}} \quad (4.48)$$

La reconstrucción de la sucesión de estados exige que se conserve el argumento que maximiza (4.48), para cada  $n$  y  $j$  lo que se puede implementar mediante un arreglo  $\psi_n(j)$ ,

calculado de forma recursiva como

$$\psi_n(j) = \arg \max_i (m_{n-1}(i)\pi_{ij}), \quad n = 2, \dots, n_\varphi - 1, \quad 1 \leq j \leq n_\vartheta$$

con valores iniciales,  $\psi_1(j) = 0$ ,  $1 \leq j \leq n_\vartheta$ .

**Ajuste de los parámetros del modelo.** Se considera que no existe una solución analítica para un modelo que maximice la probabilidad de la sucesión de observación. Esto es, dada cualquier sucesión de observación sobre un intervalo finito de tiempo, durante el entrenamiento del sistema, no hay una forma óptima para la estimación de los parámetros del modelo. Sin embargo, se puede determinar el modelo  $\mu = \{\Pi, \mathbf{P}(\mathbf{v}|\boldsymbol{\theta}), \mathbf{p}_{\theta_1}\}$ , tal que se maximiza localmente la función  $P(\varphi|\mu)$ , usando diferentes procedimientos iterativos de estimación existentes.

El primer método corresponde al método de *máxima esperanza* que implica la estimación de máxima verosimilitud y se emplea cuando la información necesaria para estimar los parámetros del modelo está incompleta, como es el caso de los modelos ocultos de Markov, en el que se conocen las sucesiones de observación, pero no se conocen las sucesiones de estados. El algoritmo se compone de dos pasos:

- *Cálculo de la esperanza* o promedio,  $\mathcal{Q}(\tilde{\mu}|\mu) = E\{\log P(\boldsymbol{\theta}, \varphi|\tilde{\mu})|\varphi, \mu\}$ ,  $\mu, \tilde{\mu} \in \mathfrak{M}$
- *Maximización*, cuando se escogen los valores de  $\tilde{\mu}$  que maximicen un funcional  $\mathcal{Q}(\tilde{\mu}|\mu)$ .

donde  $\tilde{\mu}$  representa el nuevo conjunto de parámetros después de una iteración y  $\mathfrak{M}$  es el espacio de los parámetros del modelo. El algoritmo de máxima esperanza se basa en la búsqueda de un conjunto de parámetros  $\tilde{\mu}$  que maximice  $\log P(\boldsymbol{\theta}, \varphi|\tilde{\mu})$ . Sin embargo, debido a que no se conoce la relación  $\log P(\boldsymbol{\theta}, \varphi|\tilde{\mu})$ , entonces se maximiza su valor esperado actual dados, tanto las observaciones  $\varphi$  como el modelo  $\mu$ . La función  $\mathcal{Q}(\tilde{\mu}|\mu)$  se conoce como la *función auxiliar*.

Con el fin de describir el procedimiento de estimación recursiva de los parámetros del modelo oculto de Markov, se define  $\zeta_n(i, j)$  como la probabilidad de situarse en el estado  $\vartheta_i$  en el tiempo  $n$  y en el estado  $\vartheta_j$  para el tiempo  $n + 1$ , dado el modelo y la sucesión de observación

$$\zeta_n(i, j) = P(\theta_n = \vartheta_i, \theta_{n+1} = \vartheta_j | \varphi, \tilde{\mu})$$

la cual puede ser descrita en términos de las variables de propagación en la forma,

$$\zeta_n(i, j) = \frac{\alpha_n(i)\pi_{ij}p_{j\varphi_{n+1}}\beta_{n+1}(j)}{P(\varphi|\tilde{\mu})} = \frac{\alpha_n(i)\pi_{ij}p_{j\varphi_{n+1}}\beta_{n+1}(j)}{\sum_{i=1}^{n_\vartheta} \sum_{j=1}^{n_\vartheta} \alpha_n(i)\pi_{ij}p_{j\varphi_{n+1}}\beta_{n+1}(j)}$$

donde el numerador corresponde a  $P(\theta_n = \vartheta_i, \theta_{n+1} = \vartheta_j | \varphi, \tilde{\mu})$  que dividido por  $P(\varphi|\tilde{\mu})$  es

la medida deseada de probabilidad, la cual, se relaciona con la medida (4.46),

$$\gamma_n(i) = \sum_{j=1}^{n_\vartheta} \zeta_n(i, j)$$

La probabilidad conjunta de la sucesión de observación  $\varphi$  y la sucesión de estados  $\theta$  bajo los parámetros del modelo  $\tilde{\mu}$  está dada por la ecuación (4.41) que toma la forma

$$P(\varphi, \theta | \tilde{\mu}) = p_{\theta_1}(\theta_1) \prod_{n=2}^{n_\varphi} P(\theta_{n+1} | \theta_n, \tilde{\mu}) \prod_{n=2}^{n_\varphi} P(\varphi_n | \theta_n, \tilde{\mu})$$

El logaritmo de la función de probabilidad conjunta está dado por

$$\log P(\varphi, \theta | \tilde{\mu}) = \log p_{\theta_1}(\theta_1) + \sum_{n=2}^{n_\varphi} \log P(\theta_{n+1} | \theta_n, \tilde{\mu}) + \sum_{n=2}^{n_\varphi} \log P(\varphi_n | \theta_n, \tilde{\mu}).$$

El algoritmo de máxima esperanza requiere la maximización de la función  $\mathcal{Q}(\tilde{\mu} | \mu)$ , que es el valor esperado del logaritmo de la probabilidad conjunta, donde el valor esperado se toma con respecto a la distribución anterior de los estados,  $P(\theta | \varphi, \mu)$ , dadas las observaciones. Así,

$$\begin{aligned} \mathcal{Q}(\tilde{\mu} | \mu) = \sum_{\theta} P(\theta | \varphi, \mu) \log \tilde{p}_{\theta_1}(\theta_1) + \sum_{\theta} P(\theta | \varphi, \mu) \sum_{n=2}^{N_\varphi} \log P(\theta_n | \theta_{n-1}, \tilde{\mu}) + \\ + \sum_{\theta} P(\theta | \varphi, \mu) \sum_{n=1}^{N_\varphi} \log P(\varphi_n | \theta_n, \tilde{\mu}) \end{aligned} \quad (4.49)$$

donde  $\sum_{\theta}$  denota la suma sobre todas las posibles sucesiones de estados. La ecuación (4.49) está compuesta de tres términos; el primero para los parámetros de la probabilidad inicial  $\mathbf{p}_{\theta_1}$ , el segundo para los parámetros de la matriz probabilidad de transición de estados  $\mathbf{\Pi}$  y el tercero para la matriz de probabilidad del modelo de observación  $\mathbf{P}(\mathbf{v} | \boldsymbol{\vartheta})$ ,

$$\mathcal{Q}(\tilde{\mu} | \mu) = \mathcal{Q}(\tilde{\mathbf{p}}_{\theta_1} | \mathbf{p}_{\theta_1}) + \mathcal{Q}(\tilde{\mathbf{\Pi}} | \mathbf{\Pi}) + \mathcal{Q}(\tilde{\mathbf{P}}(\mathbf{v} | \boldsymbol{\vartheta}) | \mathbf{P}(\mathbf{v} | \boldsymbol{\vartheta})) \quad (4.50)$$

Los términos en la ecuación (4.50) se pueden maximizar de forma separada. La expresión de la estimación recursiva para la densidad inicial de probabilidades  $\mathbf{p}_{\theta_1}$  puede obtenerse como

$$\mathcal{Q}(\tilde{\mathbf{p}}_{\theta_1} | \mathbf{p}_{\theta_1}) = \sum_{\theta} P(\theta | \varphi, \mu) \log \tilde{p}_{\theta_1}(\theta_1)$$

que es equivalente a la expresión

$$\mathcal{Q}(\tilde{\mathbf{p}}_{\theta_1} | \mathbf{p}_{\theta_1}) = \sum_{i=1}^{n_\vartheta} P(\theta_1 = \vartheta_i | \varphi, \mu) \log \tilde{p}_{\theta_1}(\theta_1 = \vartheta_i)$$

que, usando la ecuación (4.46), es igual a

$$\mathcal{Q}(\tilde{\mathbf{p}}_{\theta_1} | \mathbf{p}_{\theta_1}) = \sum_{i=1}^{n_{\vartheta}} \gamma_1(i) \log \tilde{p}_{\theta_1}(i)$$

La suma del factor de Lagrange  $\varrho$  en la expresión anterior, asumiendo la debida restricción,  $\sum_i p_{\theta_1}(i) = 1$  y al hacer la derivada igual a cero, resulta en

$$\frac{\partial}{\partial \tilde{p}_{\theta_1}(i)} \left( \sum_{i=1}^{n_{\vartheta}} \gamma_1(i) \log \tilde{p}_{\theta_1}(i) + \varrho \left( \sum_i \tilde{p}_{\theta_1}(i) - 1 \right) \right) = 0$$

Del desarrollo de la derivada, realizando la suma sobre  $i$  para obtener  $\varrho$  y resolviendo para  $\tilde{p}_1^{\theta}(i)$ , se obtiene la solución

$$\tilde{p}_{\theta_1}(i) = \gamma_1(i) \quad (4.51)$$

que puede calcularse usando las variables de propagación  $\alpha_1(i)$  y  $\beta_1(i)$ . En el procedimiento de promedio del algoritmo se calcula  $\gamma_1(i)$  usando el conjunto de parámetros  $\mu$ . Mientras, en el procedimiento de maximización se basa en (4.51) para obtener  $\tilde{\mu}$ .

La expresión para la actualización de los parámetros de la matriz de transición de estados  $\tilde{\Pi}$  puede encontrarse al maximizar la función,

$$\mathcal{Q}(\tilde{\Pi} | \Pi) = \sum_{\theta} P(\theta | \varphi, \mu) \sum_{n=2}^{N_{\varphi}} \log P(\theta_n | \theta_{n-1}, \tilde{\mu})$$

Sin embargo, la anterior suma puede descomponerse en la forma:

$$\begin{aligned} \mathcal{Q}(\tilde{\Pi} | \Pi) &= \sum_{i=1}^{n_{\vartheta}} \sum_{j=1}^{n_{\vartheta}} \sum_{n=2}^{n_{\varphi}} P(\theta_n = \vartheta_j, \theta_{n-1} = \vartheta_i | \mu) \log P(\theta_n | \theta_{n-1}, \tilde{\mu}) \\ &= \sum_{i=1}^{n_{\vartheta}} \sum_{j=1}^{n_{\vartheta}} \sum_{n=2}^{n_{\varphi}} \zeta_n(i, j) \log(\tilde{\pi}_{ij}) \end{aligned}$$

entonces, la maximización de la expresión anterior, con la restricción  $\sum_{j=1}^{n_{\vartheta}} \pi_{ij} = 1$ , conduce a la siguiente forma de actualización

$$\tilde{\pi}_{ij} = \frac{\sum_{n=2}^{n_{\varphi}} \zeta_n(i, j)}{\sum_{n=2}^{n_{\varphi}} \gamma_n(i)} \quad (4.52)$$

En este caso, durante la etapa de promedio, los valores de  $\zeta_n(i, j)$  y  $\gamma_n(i, j)$  se estiman usando los parámetros del modelo  $\mu$ . Mientras, para la maximización, se emplea la expresión (4.52) para calcular los nuevos parámetros  $\tilde{\mu}$ .

De otra parte, la expresión para la actualización de los parámetros del modelo de obser-

vación  $\tilde{\mathbf{P}}(\mathbf{v}|\boldsymbol{\theta})$  puede encontrarse maximizando

$$\mathcal{Q}(\tilde{\mathbf{P}}(\mathbf{v}|\boldsymbol{\theta})|\mathbf{P}(\mathbf{v}|\boldsymbol{\theta})) = \sum_{\boldsymbol{\theta}} P(\boldsymbol{\theta}|\boldsymbol{\varphi}, \mu) \sum_{n=1}^{n_{\varphi}} \log P(\varphi_n|\theta_n, \tilde{\mu})$$

De manera similar, se descompone la anterior suma,

$$\begin{aligned} \mathcal{Q}(\tilde{\mathbf{P}}(\mathbf{v}|\boldsymbol{\theta})|\mathbf{P}(\mathbf{v}|\boldsymbol{\theta})) &= \sum_{i=1}^{n_{\vartheta}} \sum_{n=1}^{n_{\varphi}} P(\theta_n = \vartheta_i|\mu) \log P(\varphi_n = v_k|\theta_n = \vartheta_i, \tilde{\mu}) \\ &= \sum_{i=1}^{n_{\vartheta}} \sum_{n=1}^{n_{\varphi}} \gamma_n(i) \log \tilde{p}_{ik} \end{aligned}$$

con lo cual, la maximización de la expresión anterior, dada la restricción

$$\sum_{k=1}^{n_{\varphi}} \tilde{P}_{ik} = 1$$

resulta en:

$$\tilde{P}_{ik} = \frac{\sum_{n=1}^{n_{\varphi}} \gamma_n(i) \delta(\varphi_n, v_k)}{\sum_{n=1}^{n_{\varphi}} \gamma_n(i)}, \quad \delta(\varphi_n, v_k) = \begin{cases} 1, & \varphi_n = v_k \\ 0, & \varphi_n \neq v_k \end{cases} \quad (4.53)$$

En este caso, la función  $\Lambda_{\varphi}(\mu) = \log P(\boldsymbol{\varphi}|\mu)$  corresponde a la respectiva relación de verosimilitud,

$$\Lambda_{\varphi}(\mu) = \mathcal{Q}(\tilde{\mu}|\mu) - \mathcal{H}(\tilde{\mu}|\mu),$$

donde  $\mathcal{H}(\tilde{\mu}|\mu) = E\{\log P(\boldsymbol{\theta}|\boldsymbol{\varphi}, \tilde{\mu})|\boldsymbol{\varphi}, \mu\}$ .

A partir de la desigualdad de Jensen [34] se obtiene  $\mathcal{H}(\tilde{\mu}|\mu) \leq \mathcal{H}(\mu|\mu)$  para cualquier  $\mu$  y  $\tilde{\mu}$  en  $\mathfrak{M}$ , cumpliéndose la igualdad, si y sólo si,  $P(\boldsymbol{\theta}|\boldsymbol{\varphi}, \mu) = P(\boldsymbol{\theta}|\boldsymbol{\varphi}, \tilde{\mu})$ . En efecto, para cualquier  $\tilde{\mu}$ ,

$$\begin{aligned} \mathcal{H}(\tilde{\mu}|\mu) - \mathcal{H}(\mu|\mu) &= E \left\{ \log \left( \frac{P(\boldsymbol{\theta}|\boldsymbol{\varphi}, \tilde{\mu})}{P(\boldsymbol{\theta}|\boldsymbol{\varphi}, \mu)} \right) \middle| \boldsymbol{\varphi}, \mu \right\} \\ &\leq \log E \left\{ \left( \frac{P(\boldsymbol{\theta}|\boldsymbol{\varphi}, \tilde{\mu})}{P(\boldsymbol{\theta}|\boldsymbol{\varphi}, \mu)} \right) \middle| \boldsymbol{\varphi}, \mu \right\} \\ &= \log \int_{\boldsymbol{\theta}} \left( \frac{P(\boldsymbol{\theta}|\boldsymbol{\varphi}, \tilde{\mu})}{P(\boldsymbol{\theta}|\boldsymbol{\varphi}, \mu)} \right) P(\boldsymbol{\theta}|\boldsymbol{\varphi}, \mu) d\boldsymbol{\theta} = \log \int_{\boldsymbol{\theta}} P(\boldsymbol{\theta}|\boldsymbol{\varphi}, \tilde{\mu}) d\boldsymbol{\theta} \\ &= 0, \end{aligned}$$

usando igualmente el hecho de la concavidad de la función logarítmica.

En los modelos en que se desarrolla el método de máxima esperanza, se asume la posi-

bilidad de evaluación de las derivadas de la relación  $\Lambda_\varphi(\mu)$  con respecto a los parámetros del modelo  $\mu$ . Como consecuencia de lo anterior, en lugar de recurrir a cualquier algoritmo específico, por ejemplo, el de máxima esperanza, directamente se pueden utilizar métodos de optimización basados en el análisis del gradiente. Concretamente, el algoritmo simple de gradiente descendente, aunque no sea el más eficiente. Sin embargo, en este método, el valor presente de los parámetros  $\tilde{\mu}_n$  se actualiza sumándole un valor de ajuste que multiplica al gradiente, en la siguiente forma recursiva:

$$\tilde{\mu}_{n+1} = \tilde{\mu}_n + \rho_n \nabla_\mu \Lambda_\varphi(\tilde{\mu}_n), \quad \rho_n > 0$$

El múltiplo  $\rho_n$  es un escalar no negativo que necesita ajustarse en cada iteración para asegurar que la sucesión  $\{\Lambda_\varphi(\tilde{\mu}_n)\}$  no sea decreciente. De cualquier manera, el algoritmo de gradiente descendente suele ser poco utilizado debido a su lenta convergencia. Para garantizar valores más rápidos de convergencias se han desarrollado diferentes algoritmos basados en propiedades de segundo orden de la función objetivo, por ejemplo, el algoritmo de Newton o el de Newton-Raphson,

$$\tilde{\mu}_{n+1} = \tilde{\mu}_n - H^{-1}(\tilde{\mu}_n) \nabla_\mu \Lambda_\varphi(\tilde{\mu}_n),$$

donde  $H^{-1}(\tilde{\mu}_n) = \nabla_\mu^2 \Lambda_\varphi(\tilde{\mu}_n)$  es la matriz Hessiana de la función objetivo. El algoritmo iterativo de Newton está basada en la siguiente aproximación de segundo orden:

$$\Lambda(\mu) \approx \Lambda(\mu') + \nabla \Lambda(\mu')(\mu - \mu') + \frac{1}{2}(\mu - \mu')^\top H(\mu')(\mu - \mu'),$$

Si la sucesión  $\{\tilde{\mu}_n\}_{n>0}$  producida por el algoritmo converge a un punto  $\tilde{\mu}^*$ , en el cual la matriz Hessiana es negativa definida, la convergencia es, por lo menos cuadrática. Por lo anterior, el procedimiento puede ser bastante eficiente. Sin embargo, una restricción seria del algoritmo de Newton se encuentra en el hecho que, excepto en el caso particular en que la función  $\Lambda(\mu)$  sea estrictamente cóncava, pueden existir regiones del espacio paramétrico en las cuales la matriz Hessiana no tenga inversa o no sea negativa semi-definida, lo que conduce a inestabilidades numéricas. Para solucionar este problema, los métodos cuasi-Newton introducen la recursión modificada

$$\tilde{\mu}_{n+1} = \tilde{\mu}_n + \rho_n W^n(\tilde{\mu}_n) \nabla_\mu \Lambda_\varphi(\tilde{\mu}_n),$$

donde  $W^n$  es una matriz de pesos que puede calcularse en cada iteración, igual que  $\rho_n$ , un factor que se usa para calcular la longitud del paso en la dirección de búsqueda. Si  $W^n$  se encuentra cercano a  $-H^{-1}(\tilde{\mu}_n)$  cuando ocurre la convergencia, el algoritmo modificado compartirá las mismas propiedades deseables del algoritmo de Newton. De otra parte, usando una matriz  $W^n$  diferente de  $-H^{-1}(\tilde{\mu}_n)$ , las restricciones numéricas asociadas con la inversión de la matriz se pueden evitar. Estos métodos construyen  $W^n$  a partir de la información del gradiente, sin la evaluación directa de la matriz Hessiana.

**Ejercicio en el CP 4.2.** Los ejemplos y ejercicios de este capítulo se realizan con base en el *toolbox*, especialmente desarrollado en MATLAB para el estudio de los modelos ocultos de Markov, que puede ser descargado en: <http://www.cs.ubc.ca/~murphyk/Software/HMM/hmm.html>. El primer paso en el entrenamiento de los modelos ocultos de Markov es su inicialización, para lo cual es necesario definir el número de estados del modelo y el número de componentes Gaussianas por estado.

```
% Parámetros del modelo de Markov:
M = 3;           % Número de componentes Gaussianas por estado
Q = 2;           % Número de estados
cov_type = 'full'; % Tipo de matriz de covarianza
```

Además, es necesario definir el tipo de matriz de covarianza que será utilizada en las componentes Gaussianas, que puede ser de los tipos: completa ('full'), diagonal ('diag') y esférica ('spherical'). A modo de ejemplo, se escoge el tipo de matriz de covarianza ('full') por cuanto permite modelar la relación entre cada una de las variables. Sin embargo, en la práctica, utilizar una matriz de covarianza completa, impide el buen desempeño del algoritmo de estimación presentándose problemas de singularidad. Por lo tanto, en muchos problemas reales es más común emplear el tipo de matriz de covarianza diagonal, en la cual se asume independencia entre los parámetros. Así mismo, se asume que se tiene un conjunto de entrenamiento y que cada trayectoria está representada por una sucesión de observaciones. La longitud de cada sucesión de observaciones se representa por T. Mientras, cada observación está compuesta por O coeficientes o parámetros. Como ejemplo concreto se generan Ns sucesiones de observaciones con parámetros:

```
% Conjunto de entrenamiento
Ns = 50; T = 10; O = 8; data = randn(O,T,Ns);
prior0 = normalise(rand(Q,1)); % vector probabilidad inicial
transmat0 = mk_stochastic(rand(Q,Q)); % Matriz de probabilidades de
                                         % transición
[mu0, Sigma0] = mixgauss_init(Q*M, data, cov_type); % media y covarianza
```

La inicialización de los parámetros del modelo (el vector probabilidad inicial, la matriz de probabilidades de transición y los parámetros de las componentes Gaussianas: vector de pesos, vector de medias y matriz de covarianza, por cada uno de los estados) debe tener en cuenta las restricciones que conservan las medidas estadísticas.

```
% Restricciones de las medidas estadísticas
mu0 = reshape(mu0, [O Q M]); Sigma0 = reshape(Sigma0, [O O Q M]);
mixmat0 = mk_stochastic(rand(Q,M)); % Vector de pesos de las mezclas Gaussianas
                                         % Entrenamiento
[LL, prior, transmat, mu, Sigma, mixmat] = ...
    mhmm_em(data, prior0, transmat0, mu0, Sigma0, mixmat0, 'max_iter',5);
% Evaluación
loglik = mhmm_logprob(data, prior, transmat, mu, Sigma, mixmat);
```

El parámetro LL corresponde al logaritmo de la verosimilitud en el entrenamiento. La evaluación de la pertenencia de una sucesión al modelo se realiza mediante el cálculo de la máxima probabilidad a posteriori de la sucesión, dado un modelo concreto.

### 4.2.3. Clasificación y comparación modelos

Hasta ahora, se han considerado las cadenas ocultas de Markov ergódicas, en las cuales a cada estado de la cadena se puede llegar en un único paso desde cualquier otro estado del modelo (estrictamente hablando, un modelo ergódico tiene la propiedad por la cual cada estado puede alcanzarse desde cualquier otro estado después de un número finito de pasos).

Así mismo, los modelos de cambio de estado también se orientan en el sentido de los valores del proceso (modelos de nacimiento y muerte). En la práctica, mayor uso han encontrado los modelos de nacimiento, cuando los coeficientes de transición de estado tienen la propiedad,  $\pi_{ij} = 0$ ,  $j < i$ , esto es, no se permiten cambios a estados con valores del proceso menores que el actual. Entonces, el vector de probabilidad de estado inicial es:

$$p_{\theta_1}(i) = \begin{cases} 0, & i \neq 1 \\ 1, & i = 1 \end{cases}$$

que implica que el proceso debe comenzar desde el estado 1 y a finalizar en el estado  $n_g$ .

Sobre los modelos de nacimiento, se suelen imponer restricciones adicionales para asegurar que no ocurran cambios mayores entre los índices de los estados. En este sentido, es usual el empleo de la restricción de la forma  $\pi_{ij} = 0$ ,  $j > i + \Delta$ . En el caso particular de los modelos de nacimiento con  $\Delta = k$ , se obtiene la siguiente matriz de transición de estado:

$$\begin{bmatrix} \pi_{11} & \pi_{12} & \cdots & \pi_{1k} & 0 & \cdots & 0 \\ 0 & \pi_{22} & \cdots & \pi_{2k} & \pi_{2(k+1)} & \cdots & 0 \\ \vdots & \vdots & & \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & 0 & 0 & \cdots & \pi_{kk} \end{bmatrix}$$

Las anteriores restricciones, impuestas sobre la orientación de la forma de cambio de los valores del modelo no afectan, en general, el procedimiento de estimación de sus parámetros.

**Densidades continuas de observación.** Las observaciones, de las cuales se estiman los parámetros de los modelos ocultos de Markov, pueden corresponder a eventos discretos, cuando se tiene un alfabeto finito a priori definido de representación del proceso y, por lo tanto, existe una FDP discreta para cada estado del modelo. En la práctica, es frecuente que las observaciones correspondan a procesos aleatorios continuos. La primera forma de análisis de estos procesos consiste en la construcción de un alfabeto via la cuantificación del proceso, cuya principal restricción está en la pérdida de información asociada a la aproximación. En este sentido, es preferible la descripción de los procesos aleatorios en los modelos ocultos de Markov mediante FDP continuas, teniendo en cuenta las restricciones necesarias para lograr que los procedimientos de estimación recursiva sean consistentes. La forma más común de representación de la FDP corresponde al modelo de mezclas finitas:

$$p_{i\varphi_n} = \sum_{m=1}^M c_{jm} \mathcal{N}_{\varphi_n}(m_{jm}, \mathbf{K}_{jm})$$



donde  $\varphi_n$  es el vector que se está modelando en el instante  $n$ ,  $c_{jm}$  es el coeficiente de mezcla para la  $m$  componente en el estado  $j$  y  $\mathcal{N}$  es cualquier densidad elípticamente simétrica o log-cóncava, con media  $\mathbf{m}_{jm}$  y matriz de covarianza  $\mathbf{K}_{jm}$  para la componente  $m$  de la mezcla en el estado  $j$ . Usualmente, en calidad de densidad  $\mathcal{N}$  se emplea la Gaussiana.

Los coeficientes de la mezcla  $c_{jm}$  deben satisfacer la siguiente restricción estocástica:  $\sum_{m=1}^M c_{jm} = 1$ ,  $1 \leq j \leq n_\vartheta$ , con  $c_{jm} \geq 0$ ,  $1 \leq j \leq n_\vartheta$  y  $1 \leq m \leq M$ , de tal manera, que la FDP sea adecuadamente normalizada:  $\int_{-\infty}^{\infty} \mathbf{p}_i \varphi d\varphi = 1$ .

Las relaciones de estimación recursiva para los coeficientes de las densidades de mezcla,  $c_{jm}$ ,  $\mathbf{m}_{jm}$ , y  $\mathbf{K}_{jk}$ , para el caso de FDP Gaussianas, empleando el algoritmo de máxima esperanza, se demuestran ser iguales a

$$\tilde{c}_{jk} = \frac{\sum_{n=1}^{n_\varphi} \gamma_n(j, k)}{\sum_{n=1}^{n_\varphi} \sum_{k=1}^M \gamma_n(j, k)} \quad (4.54a)$$

$$\tilde{\mathbf{m}}_{jk} = \frac{\sum_{n=1}^{n_\varphi} \gamma_n(j, k) \varphi_n}{\sum_{n=1}^{n_\varphi} \gamma_n(j, k)} \quad (4.54b)$$

$$\tilde{\mathbf{K}}_{jk} = \frac{\sum_{n=1}^{n_\varphi} \gamma_n(j, k) (\varphi_n - \mathbf{m}_{jk})(\varphi_n - \mathbf{m}_{jk})^\top}{\sum_{n=1}^{n_\varphi} \gamma_n(j, k)} \quad (4.54c)$$

donde  $\gamma_n(j, k)$  es la probabilidad de la observación  $\varphi_n$  dada por la componente de mezcla  $k$  del estado  $j$ , esto es:

$$\gamma_n(j, k) = \left( \frac{\alpha_n(j) \beta_n(j)}{\sum_{i=1}^{n_\vartheta} \alpha_n(i) \beta_n(i)} \right) \left( \frac{c_{jk} \mathcal{N}_{\varphi_n}(\mathbf{m}_{jk}, \mathbf{K}_{jk})}{\sum_{m=1}^M c_{jm} \mathcal{N}_{\varphi_n}(\mathbf{m}_{jm}, \mathbf{K}_{jm})} \right) \quad (4.55)$$

El término  $\gamma_n(j, k)$  generaliza la variable  $\gamma_n(i)$  en (4.46) que es el caso de una mezcla simple (un solo componente) o de una FDP discreta. La estimación de los valores  $\pi_{ij}$  se mantiene igual que para la FDP discreta de observación, (4.52). En el algoritmo de máxima esperanza, para la estimación del promedio  $\tilde{\mu}$ , se calcula  $\gamma_n(j, k)$  mediante los parámetros  $\mu$ , mientras en la maximización se usan las ecuaciones (4.54a), (4.54b) y (4.54c).

La estimación recursiva de  $c_{jk}$  es la relación entre el número esperado de veces que el sistema está en el estado  $j$  empleando la componente de mezcla  $k$ , y el número esperado de veces que está en el estado  $j$ . La estimación recursiva para el vector medio  $\mathbf{m}_{jk}$  pondera cada término del numerador de (4.54b) por cada observación, por lo que resulta un valor esperado de la porción del vector de observación que describe la componente de mezcla  $k$ . Así mismo, ocurre la interpretación para la matriz de covarianza  $\mathbf{K}_{jk}$  obtenida.

La presentación usual de las expresiones de actualización se realiza en términos de los vectores de observación  $\varphi_n$  (ecs. (4.54a), (4.54b), (4.54c)). Sin embargo, es posible expresar  $\mathbf{m}_{jk}$  y  $\mathbf{K}_{jk}$  en términos de *estadísticas suficientes promedio*, cuya estimación es independiente del número de observaciones,  $n_\varphi$ :

$$\tilde{\mathbf{m}}_{jk} = \frac{s_{\varphi,jk}}{\gamma(j,k)}, \quad (4.56a)$$

$$\tilde{\mathbf{K}}_{jk} = \frac{s_{\varphi\varphi^\top,jk}}{\gamma(j,k)} - \tilde{\mathbf{m}}_{jk}\tilde{\mathbf{m}}_{jk}^\top, \quad (4.56b)$$

donde las estadísticas suficientes se dan de la forma:

$$\gamma(j,k) = \sum_{n=1}^{n_\varphi} \gamma_n(j,k), \quad s_{\varphi,jk} = \sum_{n=1}^{n_\varphi} \gamma_n(j,k)\varphi_n, \quad s_{\varphi\varphi^\top,jk} = \sum_{n=1}^{n_\varphi} \gamma_n(j,k)\varphi_n\varphi_n^\top$$

En la estimación del promedio, el algoritmo de máxima esperanza calcula los valores  $\gamma_n(j,k)$ , las estadísticas suficientes  $\gamma(j,k)$ ,  $s_{\varphi,jk}$  y  $s_{\varphi\varphi^\top,jk}$ . Mientras, el procedimiento de maximización se basa en (4.56a) y (4.56b) para obtener  $\tilde{\boldsymbol{\mu}}$ . Desde el punto de vista de coste computacional, el reemplazo de (4.54b), (4.54c) para la maximización, resulta en un cálculo más rápido de los parámetros debido a que no se hace la suma  $\sum_n$ , por cuanto esa información está ya contenida en las estadísticas suficientes. Así, el número de operaciones en la maximización permanece constante y no depende del número de observaciones  $n_\varphi$ .

**Modelos ocultos de Markov autorregresivos.** Sea el vector de observación  $\varphi$  con componentes  $(x_0, x_1, \dots, x_{K-1})$  generado sobre un espacio con dimensión  $K$ . La FDP base para el vector de observación se considera del tipo Gaussiana, y se asume el modelo autorregresivo de orden  $p$  que relaciona los componentes de  $\varphi$  en la forma:

$$x_k = - \sum_{i=1}^p a_i x_{k-i} + e_k \quad (4.57)$$

donde  $e_k$ ,  $k = 0, \dots, K-1$  son variables aleatorias independientes Gaussianas, con media 0 y varianza  $\sigma_e^2$ ,  $a_i$ ,  $i = 1, \dots, p$  son los coeficientes de predicción o autorregresión. La FDP de  $\varphi$ , para un valor grande de  $K$ , se aproxima como,  $p(\varphi) = (2\pi\sigma_e^2)^{-K/2} \exp\{-\frac{1}{2\sigma_e^2}r(\varphi, \mathbf{a})\}$ , donde

$$r(\varphi, \mathbf{a}) = R_a(0)R(0) + 2 \sum_{i=1}^p R_a(i)R(i), \quad \mathbf{a} = [1, a_1, a_2, \dots, a_p]^\top, \quad (4.58a)$$

$$R_a(i) = \sum_{j=0}^{p-i} a_j a_{j+i}, \quad 1 \leq i \leq p \quad (4.58b)$$

$$R(i) = \sum_{j=0}^{k-i-1} x_j x_{j+i}, \quad 0 \leq i \leq p \quad (4.58c)$$

En las anteriores expresiones,  $R(i)$  es la función de correlación propia de las observaciones y  $R_a(i)$  es la función de correlación propia de los coeficientes autorregresivos. El valor residual total de predicción por ventana del proceso se determina como

$$\alpha_r = E\left\{\sum_{i=1}^k (e_i)^2\right\} = k\sigma_e^2$$

donde  $\sigma_e^2$  es la varianza por cada observación de la señal de error. Asumiendo vectores normalizados de observación,  $\tilde{\varphi} = \varphi/\sqrt{\alpha_r} = \varphi/\sqrt{k\sigma_e^2}$ , entonces la FDP toma la forma:

$$p(\tilde{\varphi}) = \left(\frac{2\pi}{k}\right)^{-k/2} \exp\left\{-\frac{k}{2}r(\tilde{\varphi}, \mathbf{a})\right\} \quad (4.59)$$

En la práctica, el factor  $k$  en (4.59) se reemplaza por la longitud efectiva de la ventana  $\tilde{k}$  que corresponde a la longitud efectiva de cada vector de observación. En el análisis de los modelos autoregresivos ocultos de Markov se asume una FDP de mezcla en la forma:

$$p_{i\varphi} = \sum_{m=1}^M c_{jm} p_{\varphi}(i; m)$$

donde cada término  $p_{\varphi}(i; m)$  es la FDP definida en (4.59) con su respectivo vector de autorregresión  $\mathbf{a}_{jm}$  o bien de forma equivalente por el vector de correlación propia,  $\mathbf{R}_{jm}$ :

$$p_{\varphi}(i; m) = \left(\frac{2\pi}{k}\right)^{-k/2} \exp\left\{-\frac{k}{2}r(\varphi, \mathbf{a}_{jm})\right\}$$

La expresión de estimación recursiva para la sucesión con correlación propia,  $R(i)$  en (4.58a), dados el estado  $j$  y la mezcla  $k$ , tiene la forma:

$$\tilde{\mathbf{R}}_{jk} = \frac{\sum_{n=1}^{N_{\varphi}} \gamma_n(j, k) \mathbf{R}_n}{\sum_{n=1}^{N_{\varphi}} \gamma_n(j, k)}$$

donde  $\mathbf{R}_n = [R_n(0), R_n(1), \dots, R_n(p)]^T$  es el vector de correlación propia definido por (4.58c) para la ventana  $n$ , mientras  $\gamma_n(j, k)$  es definida como la probabilidad de encontrarse en el estado  $j$  en el tiempo  $n$  y usando la componente de mezcla  $k$ , esto es

$$\gamma_n(j, k) = \left(\frac{\alpha_n(j)\beta_n(j)}{\sum_{j=1}^{n_{\Theta}} \alpha_n(j)\beta_n(j)}\right) \left(\frac{c_{jk}p_{\varphi_n}(j; k)}{\sum_{k=1}^M c_{jk}p_{\varphi_n}(j; k)}\right) \quad (4.60)$$

De (4.60) se observa que  $\tilde{\mathbf{R}}_{jk}$  es una suma ponderada (por la probabilidad de ocurrencia) de las funciones de correlación propia normalizadas de las ventanas en la sucesión de

observación. A partir de  $\tilde{\mathbf{R}}_{jk}$ , se resuelve el conjunto de ecuaciones normales para obtener el vector de coeficientes autorregresivos  $\tilde{\mathbf{a}}_{jk}$ , dada la  $k$  mezcla del estado  $j$ , los cuales se calculan empleando (4.58b), cerrando así el bucle de estimación.

**Estimación óptima de los parámetros del modelo.** La efectividad de los modelos ocultos de Markov está relacionada con dos aspectos básicos: el mismo proceso estocástico que cumple las restricciones del modelo y la estimación confiable de los parámetros del modelo. En este sentido, se han desarrollado diversos métodos de estimación óptima alternativos al procedimiento básico de máxima verosimilitud.

La primera aproximación consiste en generar de forma conjunta varios modelos  $\mu_\nu$ ,  $\nu = 1, \dots, V$ , de manera tal que se maximice su poder discriminante, es decir, la habilidad de cada modelo para distinguir entre sucesiones de observación generadas por el modelo correcto y aquellas generadas por modelos alternativos. El criterio básico de máxima verosimilitud consiste en el empleo de sucesiones de observación separadas  $\varphi^\nu$  a fin de derivar los parámetros de modelo para cada modelo  $\mu_\nu$ . Este criterio de optimización conduce a

$$P_\nu^* = \max_{\mu_\nu} P(\varphi^\nu | \mu_\nu)$$

Un criterio de optimización alternativo es el criterio de máximo de información mutua en el cual la información mutua promedio  $I$  entre la sucesión de observación  $\varphi^\nu$  y el conjunto completo de modelos  $\mu_C = (\mu_1, \mu_2, \dots, \mu_\nu)$  se maximiza.

La implementación del criterio de máximo de información mutua, en concordancia con la definición (2.57), es la medida:

$$I(\varphi^\nu; \mu_w) = \max_{\mu_C} \left\{ P(\varphi^\nu | \mu_\nu) - \log \sum_{w=1}^V P(\varphi^\nu | \mu_w) \right\} \quad (4.61)$$

esto es, escoger  $\mu_C$  para separar el modelo correcto  $\mu_\nu$  de todos los demás modelos alternativos sobre la sucesión de entrenamiento  $\varphi^\nu$ . Sumando (4.61) sobre todas las sucesiones de entrenamiento, se puede encontrar el conjunto de modelos con mayor separación,

$$I(\varphi^\nu; \mu_w) = \max_{\mu_C} \left\{ \sum_{\nu=1}^V \left( P(\varphi^\nu | \mu_\nu) - \log \sum_{w=1}^V P(\varphi^\nu | \mu_w) \right) \right\} \quad (4.62)$$

Hay diversas razones teóricas por las cuales no existen soluciones analíticas ni recursivas de estimación, para obtener los parámetros mediante el criterio de máximo de información mutua. En la práctica, la solución de (4.62) se lleva a cabo mediante procedimientos generalizados de optimización, por ejemplo, por el método de gradiente descendente.

El criterio de optimización del *error de clasificación mínimo* busca minimizar la probabilidad de error de clasificación a través de una representación suavizada, dada una función de pérdida. La medida del error de clasificación, que representa una medida de la distancia

entre la probabilidad de una decisión correcta y otras decisiones, se define como

$$d_\nu(\varphi) = -g_\nu(\varphi|\mu_C) + \log\left\{\frac{1}{V-1} \sum_{w, w \neq \nu} \exp(g_w(\varphi|\mu_C)\eta)\right\}^{1/\eta},$$

donde  $\eta$  es un número positivo y  $g_\nu = P(\varphi|\mu_\nu)$ ,  $\nu = 1, \dots, V$ . Esta medida de error es una función continua de los parámetros  $\mu_C$  y se debe asimilar con una regla de decisión. La medida  $d_\nu(\varphi)$  se usa como variable dentro de una función suavizada

$$f_\nu(\varphi|\mu_C) = f_\nu(d_\nu(\varphi)|\mu_C) = \frac{1}{1 + \exp(-\rho d_\nu(\varphi) + o_f)}$$

con  $o_f$  igual a cero y  $\rho \geq 1$ . Para cualquier sucesión desconocida  $\varphi^\nu$ , el desempeño del sistema se mide mediante  $f(\varphi^\nu|\mu_C) = \sum_{\nu=1}^V f_\nu(\varphi^\nu|\mu_C)$  que se usa como criterio de optimización en el algoritmo de gradiente descendente. Otros criterios de optimización de los parámetros se basan en la minimización de la razón del error empírico [35] y la minimización del error cuadrático medio.

La segunda forma de aproximación consiste en asumir que la señal no fue generada necesariamente por una fuente Markoviana, pero se ajusta a ciertas de sus restricciones, por ejemplo, tiene función de correlación positiva definida. El objetivo del procedimiento de diseño, consiste entonces en hallar los parámetros del modelo oculto que minimicen la información de discriminación o la entropía mutua entre el conjunto de FDP válidas  $Q$ , que satisfacen los valores medidos de señal, y el conjunto de FDP de los modelos ocultos,  $P_\mu$ . La información discriminante entre el conjunto  $Q$  y  $P_\mu$  se determina como:

$$D(Q|P_\mu) = \int q(\varphi) \ln(q(\varphi)/p(\varphi)) d\varphi \quad (4.63)$$

donde  $q$  y  $p$  son las funciones de densidad de probabilidad que corresponden a  $Q$  y  $P_\mu$ . Aunque las técnicas de optimización para (4.63) resultan ser relativamente complejas, se suele emplear el algoritmo de Baum adecuado para el caso de modelos ocultos de Markov.

**FDP de la duración de los estados.** En un modelo oculto de Markov, la FDP de duración inherente asociada con el estado  $\vartheta_i$  y coeficiente de transición propio  $\pi_{ii}$ , se asume de la forma  $p_i(d) = (\pi_{ii})^{d-1}(1 - \pi_{ii})$ , que para la mayoría de las señales físicas es inapropiada. En lugar, se prefiere modelar explícitamente la FDP de duración en forma analítica. En este caso, es necesario reformular las expresiones de actualización de los parámetros. Así, la variable de propagación directa  $\alpha_n(i)$  se define como,

$$\alpha_n(i) = P(\varphi_1, \varphi_2, \dots, \varphi_n, \vartheta_i \text{ termina en } n|\mu)$$

Se asume que se han visitado un total de  $r$  estados durante las primeras  $n$  observaciones y se denotan los estados como  $\theta_1, \theta_2, \dots, \theta_r$ , con duraciones asociadas con cada estado iguales a  $d_1, d_2, \dots, d_r$ . Las restricciones para  $\alpha_n(i)$  están dadas en la forma,  $\theta_n = \vartheta_i$  y  $\sum_{s=1}^r d_s = n$ .

La probabilidad  $\alpha_n(i)$  puede escribirse como

$$\alpha_n(i) = \sum_{\theta} \sum_d p_{\theta_1} p_{\theta_1}(d_1) P(\varphi_1, \varphi_2, \dots, \varphi_{d_1} | \theta_1) \pi_{\theta_1 \theta_2} p_{\theta_2}(d_2) P(\varphi_{d_1+1}, \dots, \varphi_{d_1+d_2} | \theta_2) \dots \\ \times \pi_{\theta_{r-1} \theta_r} p_{\theta_r}(d_r) P(\varphi_{d_1+d_2+\dots+d_{r-1}+1}, \dots, \varphi_n | \theta_r),$$

donde la suma se toma sobre todos los estados  $\theta$  y todas las posibles duraciones  $d$ . De manera recursiva,  $\alpha_n(i)$  se escribe como

$$\alpha_n(j) = \sum_{i=1}^{n_{\vartheta}} \sum_{d=1}^D \alpha_{n-d}(i) \pi_{ij} p_j(d) \prod_{s=n-d+1}^n p_{i\varphi_s}, \quad (4.64)$$

donde  $D$  es la duración máxima en un estado. Para inicializar el cálculo de  $\alpha_n(j)$  se usa

$$\alpha_1(i) = p_{\theta_1}(i) p_i(1) p_{i\varphi_1} \quad \alpha_2(i) = p_{\theta_1}(i) p_i(2) \prod_{s=1}^2 p_{i\varphi_s} + \sum_{\substack{j=1 \\ j \neq i}}^{n_{\vartheta}} \alpha_1(j) \pi_{ji} p_i(1) p_{i\varphi_2} \\ \alpha_3(i) = p_{\theta_1}(i) p_i(3) \prod_{s=1}^3 p_{i\varphi_s} + \sum_{d=1}^2 \sum_{\substack{j=1 \\ j \neq i}}^{n_{\vartheta}} \alpha_{3-d}(j) \pi_{ji} p_i(d) \prod_{s=4-d}^3 p_{i\varphi_s}$$

De igual forma, se calcula  $\alpha_4(i)$  hasta  $\alpha_D(i)$ . Finalmente, se usa (4.64) para  $n > D$ . La probabilidad  $P(\varphi|\mu)$  se calcula usando (4.44). A fin de dar expresiones de actualización en un modelo oculto de Markov de duración variable, se definen las variables:

$$\alpha_n^*(i) = P(\varphi_1, \varphi_2, \dots, \varphi_n, \vartheta_i \text{ empieza en } n+1 | \mu) \\ \beta_n(i) = P(\varphi_{n+1}, \varphi_{n+2}, \dots, \varphi_{n_{\varphi}} | \vartheta_i \text{ termina en } n, \mu) \\ \beta_n^*(i) = P(\varphi_{n+1}, \varphi_{n+2}, \dots, \varphi_{n_{\varphi}} | \vartheta_i \text{ empieza en } n+1, \mu)$$

Las relaciones entre  $\alpha, \alpha^*, \beta$  y  $\beta^*$  son

$$\alpha_n^*(i) = \sum_{j=1}^{n_{\vartheta}} \alpha_n(i) \pi_{ij}, \quad \alpha_n(i) = \sum_{d=1}^D \alpha_{n-d}^*(i) p_i(d) \prod_{s=n-d+1}^n p_{i\varphi_s} \\ \beta_n(i) = \sum_{j=1}^{n_{\vartheta}} \pi_{ij} \beta_n^*(j) \quad \beta_n^*(i) = \sum_{d=1}^D \beta_{n+d}(i) p_i(d) \prod_{s=n+1}^{n+d} p_{i\varphi_s}$$

Con base en las definiciones y relaciones anteriores, las expresiones de actualización para un modelo variable se determinan como:

$$\tilde{p}_{\theta_1}(i) = \frac{p_{\theta_1}(i) \beta_1^*(i)}{P(\varphi|\mu)}, \quad \tilde{\pi}_{ij} = \frac{\sum_{n=1}^{n_{\varphi}} \alpha_n(i) \pi_{ij} \beta_n^*(j)}{\sum_{j=1}^{n_{\vartheta}} \sum_{n=1}^{n_{\varphi}} \alpha_n(i) \pi_{ij} \beta_n^*(j)}$$

ademàs

$$\begin{aligned}
 \tilde{P}_{ik} &= \frac{\sum_{n=1}^{n_\varphi} \left( \sum_{d < n} \alpha_d^*(i) \beta_d^*(i) - \sum_{d < n} \alpha_d(i) \beta_d(i) \right) \delta(\varphi_n, v_k)}{\sum_{k=1}^{n_\nu} \sum_{n=1}^{n_\varphi} \left( \sum_{d < n} \alpha_d^*(i) \beta_d^*(i) - \sum_{d < n} \alpha_d(i) \beta_d(i) \right) \delta(\varphi_n, v_k)} \\
 \tilde{p}_i(d) &= \frac{\sum_{n=1}^{n_\varphi} \alpha_n^*(i) p_i(d) \beta_{n+d}(i) \prod_{s=n+1}^{n+d} p_{i\varphi_s}}{\sum_{d=1}^D \sum_{n=1}^{n_\varphi} \alpha_n^*(i) p_i(d) \beta_{n+d}(i) \prod_{s=n+1}^{n+d} p_{i\varphi_s}} \quad (4.68)
 \end{aligned}$$

Al introducir la FDP para la duración de los estados, para algunos problemas, la calidad del modelo mejora significativamente. Sin embargo, aparecen algunos inconvenientes tales como un incremento considerable en la carga computacional y la necesidad de estimar un número adicional de parámetros,  $D$ , por estado, lo cual requiere un mayor número de observaciones. Una posible solución consiste en usar una FDP de duración de estado paramétrica, en lugar de la no paramétrica de la ec. (4.68). En [36], se propone la  $\Gamma$ -densidad para representar la FDP de duración de los estados,  $p_i(d) = \frac{\eta_i^{\nu_i} d^{\nu_i-1} e^{-\eta_i d}}{\Gamma(\nu_i)}$ ,  $d > 0$ , con parámetros  $\nu_i$  y  $\eta_i$ , media  $\nu_i/\eta_i$  y varianza  $\nu_i/\eta_i^2$ . Las expresiones de actualización para  $\nu_i$  y  $\eta_i$ , están dadas por

$$\tilde{\eta}_j = \frac{\nu_j \sum_{n=1}^{n_\varphi} \alpha_n(j) \beta_n(j)}{\sum_{n=1}^{n_\varphi} \sum_{d \leq n} d \sum_{\substack{i=1 \\ i \neq j}}^{n_\vartheta} \alpha_{(n-d)}(i) \pi_{ij} p_j(d) \prod_{s=1}^d p_{j\varphi_{n-d+s}} \beta_n(j)} \quad (4.69)$$

$$F(\tilde{\nu}_j) = \frac{\sum_{n=1}^{n_\varphi} \log(\eta_j d) \sum_{\substack{i=1 \\ i \neq j}}^{n_\vartheta} \alpha_{(n-d)}(i) \pi_{ij} p_j(d) \prod_{s=1}^d p_{j\varphi_{n-d+s}} \beta_n(j)}{\sum_{n=1}^{n_\varphi} \alpha_n(j) \beta_n(j)}, \quad (4.70)$$

donde  $F(\cdot)$  es la función digamma, la cual es diferenciable continuamente y puede obtenerse en forma de serie de potencia, entonces, (4.70) se resuelve numéricamente para  $\tilde{\nu}_j$ .

En las expresiones (4.69) y (4.70),  $\alpha_n(j)$  y  $\beta_n(j)$  están dados por

$$\begin{aligned}
 \alpha_n(j) &= \sum_{d \leq n} \sum_{\substack{i=1 \\ i \neq j}}^{n_\vartheta} \alpha_{(n-d)}(i) \pi_{ij} p_j(d) \prod_{s=1}^d p_{j\varphi_{n-d+s}}, \\
 \beta_n(j) &= \sum_{d \leq n} \sum_{\substack{j=1 \\ j \neq i}}^{n_\vartheta} \pi_{ij} p_j(d) \prod_{s=1}^d p_{j\varphi_{n+s}} \beta_{n+d}(j)
 \end{aligned}$$

**Ejercicio en el CP 4.3.** Sean dos procesos aleatorios, cada uno de los cuales corresponde al resultado de pasar RBG por un sistema lineal invariante en el tiempo con diferente función de transferencia del tipo respuesta a impulso infinita, (1.56):

$$H_0(z) = \frac{z+1}{2z}, \quad H_1(z) = \frac{1}{2}(1-z^{-1})$$

El proceso aleatorio generado, tiene un número de sucesiones  $N_s$ , y cada sucesión contiene un número de observaciones  $T$ .

```
% Generación del proceso RBG
data = randn(0*T*Ns,1);
% Definición del sistema lineal invariante en el tiempo H0
% Paso del RBG a través de H0
num0 = [1 1]; den0 = [2]; data0tem = filter(num0,den0,data);
% Definición del sistema lineal invariante en el tiempo H1
% Paso del RBG a través de H1
num1 = [1 -1]; den1 = [2 0]; data1tem = filter(num1,den1,data);
% Acondicionamiento de las sucesiones para ser utilizadas en el
% entrenamiento de los modelos ocultos de Markov
for i=1:length(data)/(0*T)
    data0T(i,:) = data0tem((i-1)*0*T+1:i*0*T)';
    data1T(i,:) = data1tem((i-1)*0*T+1:i*0*T)';
    for j = 1:size(data0T,2)/T
        data0(:,j,i) = data0T(i,(j-1)*T+1:j*T)';
        data1(:,j,i) = data1T(i,(j-1)*T+1:j*T)';
    end
end
```

A partir de los procesos generados en el paso anterior, es posible inicializar los parámetros de los modelos ocultos de Markov, de igual forma como se realizó en el ejercicio PC 4.2.

```
% Inicialización de los parámetros de los modelos ocultos de Markov para la clase 0
prior0 = normalise(rand(Q,1)); transmat0 = mk_stochastic(rand(Q,Q));
[mu0,Sigma0]=mixgauss_init(Q*M,data0, cov_type);
% Restricciones de las medidas estadísticas
mu0 = reshape(mu0, [0 Q M]); Sigma0 = reshape(Sigma0, [0 0 Q M]);
mixmat0 = mk_stochastic(rand(Q,M));
% Entrenamiento del modelo para la clase 0
[LL, priorC0, transmatC0, muC1, SigmaC0, mixmatC0] = mhmm_em(data0, prior0, ...
transmat0,mu0,Sigma0, mixmat0,'max_iter',5);
% Inicialización de los parámetros de los modelos ocultos de Markov para la clase 1
[mu0, Sigma0] = mixgauss_init(Q*M, data1, cov_type);
% Restricciones de las medidas estadísticas
mu0 = reshape(mu0, [0 Q M]); Sigma0 = reshape(Sigma0, [0 0 Q M]);
% Entrenamiento del modelo para la clase 1
[LL, priorC1, transmatC1, muC1, SigmaC1, mixmatC1] = mhmm_em(data0, prior0, ...
transmat0,mu0,Sigma0, mixmat0, 'max_iter',5);
```

Si se tiene una trayectoria nueva, se calcula la probabilidad a posteriori de que cada uno de los modelos genere la sucesión dada. La sucesión es asignada a la clase del modelo que genere la mayor probabilidad.

```
%se asume que la muestra a reconocer es de la clase 1
data3 = data1; loglik0 = mhmm_logprob(data3, priorC0, transmatC0, muC0, SigmaC0, ...
mixmatC0);
loglik1 = mhmm_logprob(data3, priorC1, transmatC1, muC1, SigmaC1, ...
mixmatC1);
if loglik1 > loglik0
    class = 1;
else
    class = 0;
end
```



**Comparación de los modelos de ocultos de Markov.** Dados dos procesos aleatorios de Markov de primer orden  $\mu_1$  y  $\mu_2$  se debe hallar una medida de similitud o distancia entre los mismos,  $d(\mu_1, \mu_2)$ , con el fin de medir su equivalencia estadística. En la práctica, se han propuesto diversas medidas de distancia. La primer medida propuesta corresponde a una forma generalizada de la distancia euclídea entre las matrices de probabilidad estado-observación, que se define como:

$$d(\mu_1, \mu_2) = \sqrt{\frac{1}{n_\vartheta} \sum_{i=1}^{n_\vartheta} \sum_{k=1}^{n_v} \|P_{ik}^{(1)} - P_{ik}^{(2)}\|^2}, \quad (4.71)$$

Un caso más general de esta medida se realiza encontrando el estado del segundo proceso que minimiza la diferencia entre las probabilidades del los modelos,

$$d(\mu_1, \mu_2) = \left\{ \frac{1}{n_\vartheta n_v} \sum_{i=1}^{n_\vartheta} \sum_{k=1}^{n_v} \left( P_{ik}^{(1)} - P_{c(i)k}^{(2)} \right)^2 \right\}^{1/2} \quad (4.72)$$

donde  $c(i)$  es la permutación de los estados que minimiza (4.72). Las medidas (4.71) y (4.72) son inadecuadas, dado que no toman en cuenta la estructura temporal representada en la cadena de Markov, por lo que podría darse el caso en el cual es posible encontrar un par de modelos ocultos de Markov, con una distancia entre sí que tienda a cero, pero con medidas respectivas de probabilidad,  $P_\mu$  y  $P_{\mu'}$ , completamente diferentes.

Una medida alterna, que define la distancia de *Kullback - Leibler* entre las probabilidades (4.44),  $P(\cdot|\mu_i)$ , está dada por

$$d(\mu_1, \mu_2) = \frac{1}{2} (\log(P_{11}P_{22}) - \log(P_{12}P_{21})), \quad (4.73)$$

donde

$$P_{ij} = P(\varphi_i|\mu_j)^{1/n_{\varphi_i}}, \quad (4.74)$$

siendo  $n_{\varphi_i}$  la longitud de la sucesión  $\varphi_i$  generada de forma estocástica a partir del modelo  $\mu_i$ . La medida  $d(\mu_1, \mu_2)$  está determinada unívocamente sólo en el límite, cuando  $n_{\varphi_i} \rightarrow \infty$ . Así mismo, se puede demostrar que, si  $P_{ii}$  es un máximo global, la distancia de *Kullback-Leibler* tiene las siguientes propiedades [37]:

1.  $d(\mu_1, \mu_2) = d(\mu_2, \mu_1)$
2.  $d(\mu_1, \mu_2) \geq 0$
3.  $d(\mu_1, \mu_2) = 0$  si  $\mu_1 \sim \mu_2$  ó  $\varphi_1 = \varphi_2$

Una de las desventajas que tiene este tipo de distancia se encuentra en la necesidad de realizar la simulación, por ejemplo por Monte Carlo, para generar las sucesiones  $\hat{\varphi}_i$ , lo cual eleva el costo computacional. En [38], se propone una cota superior para la distancia de Kullback Leibler, que hace innecesario el procedimiento de simulación.

Otras distancias entre modelos ocultos de Markov se definen sobre el concepto de *probabilidad de co-emisión* [39],  $\sum_{\varphi} \mu_1(\varphi) \mu_2(\varphi)$ , que es la probabilidad de que independientemente los modelos generen la misma sucesión. Esta probabilidad requiere que el modelo del proceso de Markov tenga la estructura de un modelo de nacimiento. El cálculo de la probabilidad de co-emisión implica la construcción de una tabla  $\mathbf{A}$  indexada por los estados de ambos modelos ocultos de Markov, tal que la entrada  $P_{\mathbf{A}}(\theta_{\mu_1}, \theta_{\mu_2})$  de la tabla (donde  $\theta_{\mu_i}$  es un estado de  $\mu_i$ ,  $i = 1, 2$ ) representa la probabilidad de encontrarse en el estado  $\theta_{\mu_1}$  en  $\mu_1$  y  $\theta_{\mu_2}$  en  $\mu_2$ , además de haber generado independientemente sucesiones idénticas de observación en las rutas de estados, que conducen a  $\theta_{\mu_1}$  y a  $\theta_{\mu_2}$ . El valor de la tabla correspondiente a los dos estados finales de cada modelo es la probabilidad de co-emisión.

Basados en la probabilidad de co-emisión de dos modelos dados,  $P_{\mathbf{A}}(\mu_1, \mu_2)$ , se definen las siguientes dos métricas:

$$d_{ang}(\mu_1, \mu_2) = \arccos(P_{\mathbf{A}}(\mu_1, \mu_2) / \sqrt{P_{\mathbf{A}}(\mu_1, \mu_1) P_{\mathbf{A}}(\mu_2, \mu_2)})$$

$$d_{dif}(\mu_1, \mu_2) = \sqrt{P_{\mathbf{A}}(\mu_1, \mu_1) + P_{\mathbf{A}}(\mu_2, \mu_2) - 2P_{\mathbf{A}}(\mu_1, \mu_2)}$$

Otra versión de métrica se conoce como la *medida de similitud* que mide la semejanza entre dos trayectorias estocásticas con dimensión múltiple en correspondencia con los modelos ocultos de Markov dados,

$$d(\varphi_1, \varphi_2) = \sqrt{\frac{P_{21}P_{12}}{P_{11}P_{22}}}$$

donde,  $P_{ij} = P(\varphi_i | \mu_j)^{n_{\varphi_i}}$ , representa la probabilidad de la sucesión de observación  $\varphi_i$  dada por el modelo  $\mu_j$ , normalizado con respecto a  $n_{\varphi_i}$ , donde  $n_{\varphi_i}$  es la longitud de la sucesión  $\varphi_i$ . Se puede demostrar [40] que si  $P_{ii}$  es un máximo global, la medida de similitud tiene las siguientes propiedades:

1.  $d(\varphi_1, \varphi_2) = d(\varphi_2, \varphi_1)$
2.  $0 < d(\varphi_1, \varphi_2) \leq 1$
3.  $d(\varphi_1, \varphi_2) = 1$  si  $\mu_1 \sim \mu_2$  ó  $\varphi_1 = \varphi_2$

En algunas ocasiones es más conveniente representar la similitud entre dos modelos de Markov a través de una medida de distancia en lugar de una medida de similitud. Dada la medida de similitud  $d(\varphi_1, \varphi_2)$ , la medida de distancia se puede obtener a partir de

$$d(\varphi_1, \varphi_2) = -\log d(\varphi_1, \varphi_2) \quad (4.75)$$

tal que se cumple  $d(\varphi_1, \varphi_2) = d(\varphi_2, \varphi_1)$ ,  $d(\varphi_1, \varphi_2) \geq 0$  y  $d(\varphi_1, \varphi_2) = 1$  si  $\mu_1 \sim \mu_2$  ó  $\varphi_1 = \varphi_2$ .

A diferencia de las sucesiones de observación  $\varphi_i$ , las sucesiones de observación  $\widehat{\varphi}_i$  de (4.74) no son únicas debido a que son generadas de forma estocástica a partir de  $\widehat{\mu}_i$ .

Aunque de forma general  $d(\hat{\mu}_1, \hat{\mu}_2)$  (ecuación (4.73)) y  $d(\varphi_1, \varphi_2)$  (ecuación (4.75)) no son equivalentes, bajo ciertas presunciones las dos nociones (distancia entre los modelos ocultos y distancia entre sucesiones de observación) convergen a una equivalencia. Específicamente, se tiene que,  $d(\varphi_1, \varphi_2) = d(\hat{\mu}_1, \hat{\mu}_2)$  si y sólo si

1.  $\mu_1 \sim \hat{\mu}_1$  y  $\mu_2 \sim \hat{\mu}_2$
2.  $P_{11}$  y  $P_{22}$  son máximos globales.
3.  $\hat{n}_{\varphi_i} \rightarrow \infty$

**Ejercicio en el CP 4.4.** Dados dos modelos ocultos de Markov, calcular la distancia de Kullback-Leibler entre los modelos. Para realizar este ejercicio, serán tomados los modelos que fueron entrenados a partir de procesos aleatorios diferentes, en el ejercicio PC 4.3. El modelo generado a partir de las sucesiones de observaciones 0, es llamado modelo clase  $C_0$ . De igual forma, el modelo generado a partir de las sucesiones de observaciones 1, es llamado modelo clase  $C_1$ . Esta notación se mantiene para cada uno de los parámetros del modelo.

En primer lugar es necesario definir la longitud de las sucesiones de observación. Debe tenerse en cuenta, que la medida de distancia Kullback-leibler está definida para una longitud de la sucesión de observaciones tendiente a infinito, por lo tanto, es necesario determinar una longitud lo suficientemente grande. En este caso, se asume una longitud  $T = 10000$ .

```
% Implementación de la distancia de Kullback - Leibler
% Definición de la longitud de las sucesiones de observación.
T = 10000;          %la longitud de las secuencias debe tender a inf
nseq = 1;           % número de secuencias
```

Después de definir la longitud de las sucesiones y el número de sucesiones, es necesario generar las sucesiones de observación a partir de los modelos dados y subsecuentemente se calculan las probabilidades definidas en la ecuación (4.74).

```
% Generación de las sucesiones a partir de los modelos dados
% Generación de sucesiones a partir del modelo C0
[obs0, hidden0] = mhmm_sample(T, nseq, priorC0, transmatC0, muC0, SigmaC0, mixmatC0);
% Generación de sucesiones a partir del modelo C1
[obs1, hidden1] = mhmm_sample(T, nseq, priorC1, transmatC1, muC1, SigmaC1, mixmatC1);
%cálculo de las probabilidades Pij
loglik00 = mhmm_logprob(obs0, priorC0, transmatC0, muC0, SigmaC0, mixmatC0);
loglik01 = mhmm_logprob(obs0, priorC1, transmatC1, muC1, SigmaC1, mixmatC1);
loglik10 = mhmm_logprob(obs1, priorC0, transmatC0, muC0, SigmaC0, mixmatC0);
loglik11 = mhmm_logprob(obs1, priorC1, transmatC1, muC1, SigmaC1, mixmatC1);
```

Según la definición (4.74), las medidas de probabilidad  $P_{ij}$  están normalizadas con respecto a la longitud de las sucesiones de observación, por lo tanto, antes de calcular la medida de distancia se debe realizar la normalización y posteriormente aplicar la definición (4.73), para así calcular la medida de distancia.

```
%Normalización
loglik00 = loglik00/T; % Se hace lo mismo para las demás f. de verosimilitud
% Cálculo de la medida de distancia
d12=loglik11-loglik12; d21=loglik22-loglik21;
% Cálculo de la distancia simétrica
Ds=(d12+d21)/2;
```

## Problemas

**Ejercicio en el CP 4.5.** Realice un análisis comparativo de diferentes modelos ocultos de Markov entrenados con el mismo conjunto de datos, variando el número de estados y de mezclas Gaussianas en el modelo.

**Ejercicio en el CP 4.6.** Calcular la secuencia de estados más probable para una sucesión y un modelo dado, empleando las funciones:

```
obslik = mixgauss_prob(Secuencia, mu, Sigma, mixmat);  
[path,loglik] = viterbi_path(prior, transmat,obslik, 0);
```

El parámetro loglik, en este caso, contiene la probabilidad conjunta de la sucesión de observación y la sucesión de estados. Realizar el análisis comparativo de ésta probabilidad conjunta y la probabilidad máximo a posteriori, para diferentes modelos variando el número de estados y de mezclas en el modelo.

**Ejercicio en el CP 4.7.** Tomar como base la implementación de la distancia de Kullback- Leibler desarrollada en ejemplo 4.4, e implementar la medida de distancia por similitud (ecuación (4.75)), compare el comportamiento de ambas medidas de distancia para un número diferente de estados y de mezclas en el modelo.

# Capítulo 5

## Transformación de procesos aleatorios

En general, el paso de procesos aleatorios por circuitos lineales consiste en el análisis de las propiedades de aleatoriedad de los procesos de entrada y salida del sistema, dada su forma de transformación. Por regla, la solución general es difícil de alcanzar, por lo tanto, ésta se simplifica hasta hallar las funciones de los momentos para el caso no estacionario, o de los valores de los momentos y la función de correlación (o su respectiva densidad espectral de potencia) para el caso estacionario.

### 5.1. Paso de señales aleatorias por sistemas lineales

En concordancia con la descripción de sistemas lineales hecha en el numeral §1.2.3, se presentan los siguientes tres métodos de análisis.

#### 5.1.1. Análisis en el tiempo

La forma directa de análisis es la convolución (1.39), que relaciona la señal aleatoria de salida  $\eta(t)$  con la entrada  $\xi(t)$ , a través de la respuesta a impulso del sistema  $h(t)$ :

$$\eta(t) = \int_0^{\infty} \xi(t - \tau) h(\tau) d\tau \quad (5.1)$$

El límite superior de la integral en (5.1) es escogido igual al valor  $\tau \rightarrow \infty$ , a fin de analizar el comportamiento asintótico del sistema en régimen de estado estable, antes que hacerlo igual a  $t$ , que corresponde al régimen transitorio.

El operador de promedio sobre ambos términos de (5.1) se generaliza en la forma:

$$E\left\{\int_T \xi(t) g(t) dt\right\} = \int_T E\{\xi(t)\} g(t) dt$$

donde  $g(t)$  es una función no aleatoria que varía en el tiempo.

El intercambio en la operación de integración con el operador de promedio, implica la simplificación del análisis; intercambio que se realiza cuando se cumplen las condiciones:

$$(a). \int_T E \{ |\xi(t)| \} g(t) dt < \infty$$

Al asumir la estacionariedad de los procesos de entrada, cuando los operadores de promedio son invariantes en el tiempo (ver (3.1)), entonces, se consideran una constante  $k$  que sale fuera de la integral. En calidad de función  $g(t)$  representativa del sistema se emplea la respuesta a impulso  $h(t)$ , por lo que la restricción, que corresponde a la condición de estabilidad del sistema, se resume a  $k \int_T |h(t)| dt < \infty$ .

$$(b). \xi(t) \in T, \text{ aunque se puede tener que } T \rightarrow \infty.$$

**Valor medio de salida.** Para un sistema lineal e invariante en el tiempo, cuando a su entrada se tiene una señal estacionaria, teniendo en cuenta (5.1), está dado por

$$\begin{aligned} E \{ \eta(t) \} &= E \left\{ \int_0^\infty \xi(t-\lambda) h(\lambda) d\lambda \right\} = \int_0^\infty E \{ \xi(t-\lambda) \} h(\lambda) d\lambda \\ &= m_{1\xi} \int_0^\infty h(\lambda) d\lambda \end{aligned} \quad (5.2)$$

Sin embargo, la integral de la respuesta a impulso en (5.2) se puede representar de forma alterna como la TF para el caso único de análisis de la componente  $\omega = 0$ , esto es,

$$\begin{aligned} E \{ \eta(t) \} &= m_{1\xi} \int_0^\infty h(\lambda) d\lambda = m_{1\xi} \int_0^\infty h(\lambda) e^{-j\omega\lambda} d\lambda \Big|_{\omega=0} \\ &= m_{1\xi} H(0) \end{aligned} \quad (5.3)$$

La expresión (5.3) es evidente, porque muestra que el primer momento, al ser una constante, afecta solamente la componente  $\omega = 0$  de la función de transferencia del sistema. Sin embargo, un proceso aleatorio a la salida de un circuito lineal, activado por una entrada a partir de  $t = 0$ , se expresa por la integral (1.39), que se diferencia de la (5.1) por su límite superior. Por lo tanto, la expresión (5.2) para el valor medio se corrige en la forma:

$$\begin{aligned} E \{ \eta(t) \} &= m_{1\xi} \int_{-\infty}^t h(\lambda) d\lambda = m_{1\xi} a(t) \\ &= m_{1\eta}(t) \end{aligned} \quad (5.4)$$

El valor medio para cualquier sistema lineal resulta directamente proporcional a la función de transición del sistema  $a(t)$ , que a su vez depende del tiempo, entonces, el proceso a la salida es no estacionario, mientras el sistema no esté en régimen de estado estable.

**Ejemplo 5.1.** En el circuito  $RC$  del ejemplo 1.3, se tiene una señal aleatoria  $\xi(t)$  de voltaje, definida a partir del momento  $t = 0$  que corresponde a RBG con valor medio  $m_{1\xi}$ . Determinar el cambio del valor medio en el tiempo del voltaje de salida  $\eta(t)$

La ecuación diferencial del circuito  $RC$  de la Figura 1.4(a) tiene la forma:

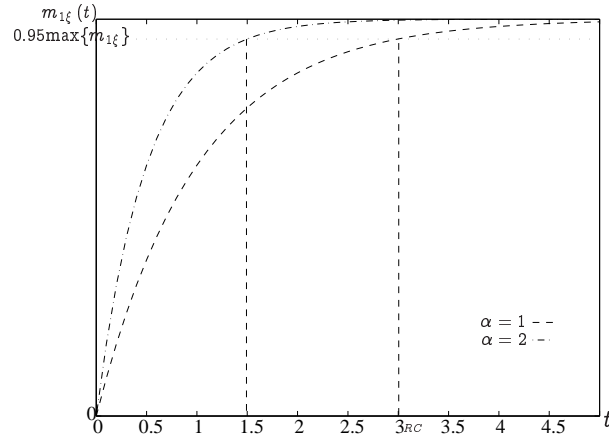
$$\frac{d\eta}{dt} + \alpha\eta(t) = \alpha\xi(t), \quad \alpha = 1/RC$$

Considerando que el capacitor en el momento  $t = 0$  está descargado, y por ende,  $\eta(0) = 0$ , la anterior ecuación diferencial tiene la siguiente solución:

$$\eta(t) = \alpha e^{-\alpha t} \int_0^t e^{\alpha\lambda} \xi(\lambda) d\lambda$$

Por lo que el valor medio del voltaje de salida  $\eta(t)$  del circuito  $RC$  es igual a:

$$E\{\eta(t)\} = \alpha e^{-\alpha t} \int_0^t e^{\alpha\lambda} E\{\xi(\lambda)\} d\lambda = \alpha e^{-\alpha t} \int_0^t e^{\alpha\lambda} m_{1\xi} d\lambda = m_{1\xi} \alpha e^{-\alpha t} (e^{\alpha t} - 1)$$



**Figura 5.1.** Cambio del valor medio en el tiempo

De esta manera, el valor medio a la salida cambia en el tiempo de la siguiente forma:

$$m_{1\eta}(t) = m_{1\xi} \alpha e^{-\alpha t} (e^{\alpha t} - 1)$$

La Figura 5.1 representa el cambio del valor medio (5.4) de la señal aleatoria de salida  $\eta(t)$  en el tiempo (para dos valores diferentes de  $\alpha$ ), que corresponde a un intervalo de no estacionariedad del proceso. Este intervalo es transitorio y su existencia, siendo proporcional al factor de inercia  $\alpha = RC$  del sistema, se considera insignificante con relación al intervalo total de análisis del proceso.

Cuando a la entrada del sistema lineal se tiene un proceso no estacionario, determinado a partir del momento  $t = 0$ , el valor medio del proceso a la salida, dado por la relación (5.2), es igual a:

$$E\{\eta(t)\} = \int_0^t h(\lambda) m_{1\xi}(t - \lambda) d\lambda = h(t) * m_{1\xi}(t) \quad (5.5)$$

La expresión (5.5) coincide con el resultado obtenido en (3.6a), para el caso de transformación lineal de señales aleatorias.

En general, el análisis en estado transitorio del sistema lineal se puede realizar para cada uno de los momentos, tanto en el caso estacionario, como en el no estacionario. Sin embargo, debido a la dificultad en la solución del caso transitorio, es frecuente que el análisis se limite al régimen de estado estable.

**Valor cuadrático medio de salida.** El análisis es similar al caso anterior del valor medio de la señal aleatoria de salida, aunque, aquí es preferible introducir dos variables de integración, a fin de representar la multiplicación interna del operador de promedio en forma de una integral doble, de la siguiente forma:

$$\begin{aligned} E\{\eta^2(t)\} &= E\left\{\int_0^\infty \xi(t - \lambda_1) h(\lambda_1) d\lambda_1 \int_0^\infty \xi(t - \lambda_2) h(\lambda_2) d\lambda_2\right\} \\ &= E\left\{\int_0^\infty d\lambda_1 \int_0^\infty \xi(t - \lambda_1) \xi(t - \lambda_2) h(\lambda_1) h(\lambda_2) d\lambda_2\right\} \\ &= \int_0^\infty d\lambda_1 \int_0^\infty E\{\xi(t - \lambda_1) \xi(t - \lambda_2)\} h(\lambda_1) h(\lambda_2) d\lambda_2 \end{aligned} \quad (5.6)$$

La operación de promedio dentro de la integral corresponde a la función de correlación de la señal aleatoria de entrada, evaluada en los siguientes puntos:

$$\begin{aligned} E\{\xi(t - \lambda_1) \xi(t - \lambda_2)\} &= R_\xi(t - \lambda_1 - t + \lambda_2) \\ &= R_\xi(\lambda_2 - \lambda_1) \end{aligned}$$

Por lo que la expresión (5.6) se formula en la siguiente forma:

$$E\{\eta^2(t)\} = \int_0^\infty d\lambda_1 \int_0^\infty R_\xi(\lambda_2 - \lambda_1) h(\lambda_1) h(\lambda_2) d\lambda_2 \quad (5.7)$$

---

**Ejemplo 5.2.** En el circuito  $RC$  del ejemplo 1.3, se tiene a la entrada una señal aleatoria  $\xi(t)$  de voltaje, que corresponde a RBG. Determinar el valor cuadrático medio del voltaje de salida  $\eta(t)$ .



El valor cuadrático medio de salida determinado por (5.7), teniendo en cuenta el valor de la función de correlación del RBG (3.34), es igual a,

$$\begin{aligned} E\{\eta^2(t)\} &= \int_0^\infty d\lambda_1 \int_0^\infty \frac{N_0}{2} \delta(\lambda_2 - \lambda_1) h(\lambda_1) h(\lambda_2) d\lambda_2 = \frac{N_0}{2} \int_0^\infty h^2(\lambda) d\lambda \\ &= \frac{N_0}{2} \int_0^\infty (\alpha e^{-\alpha\lambda})^2 d\lambda = \frac{N_0 \alpha^2}{2} \int_0^\infty e^{-2\alpha\lambda} d\lambda = \frac{N_0 \alpha^2}{2} \left. \frac{e^{-2\alpha\lambda}}{-2\alpha} \right|_0^\infty \\ &= \frac{N_0}{2} \frac{\alpha}{2} \end{aligned}$$

En consideración, cuando en vez de asumir un ruido con varianza infinita, se asume un ruido con correlación nula, valor medio cero, pero varianza finita,  $\sigma_\xi^2 = D < \infty$ , el proceso a la salida del dispositivo lineal es unívocamente cero,

$$\eta(t) = \int_0^t h(\tau) \xi(t - \tau) d\tau \equiv 0$$

en la medida en que los momentos (5.5) y (5.7) son iguales a 0.

Cabe anotar, que la resolución de este ejemplo es relativamente simple, pero solo si el proceso a la entrada es del RBG, en caso contrario, cuando el proceso es otra naturaleza, la resolución de la integral doble (5.7) presenta serias dificultades.

**Función de correlación de salida.** En este caso, se puede tomar el resultado obtenido en el análisis de la transformación de señales aleatorias con operadores lineales. Particularmente, asumiendo la estacionariedad del proceso de entrada de (3.6b), se obtiene:

$$R_\eta(\tau) = \int_0^\infty d\lambda_1 \int_0^\infty R_\xi(\lambda_2 - \lambda_1 - \tau) h(\lambda_1) h(\lambda_2) d\lambda_2 \quad (5.8)$$

La varianza del proceso a la salida resulta de hacer  $\tau = 0$  en la expresión (5.8), inclusive para el caso de análisis de regimen transitorio

$$\sigma_\eta^2(t) = \int_0^t d\lambda_1 \int_0^t R_\xi(\lambda_2 - \lambda_1) h(\lambda_1) h(\lambda_2) d\lambda_2$$

**Ejemplo 5.3.** En el circuito  $RC$  del ejemplo 1.3, se tiene una señal aleatoria  $\xi(t)$  de voltaje, definida a partir del momento  $t = 0$  que corresponde a RBG. Determinar el cambio de la función de correlación en el tiempo del voltaje de salida  $\eta(t)$ .

La expresión (5.8), sustituyendo la función de correlación del RBG (3.34), toma la forma:

$$\begin{aligned} R_\eta(\tau) &= \int_0^\infty d\lambda_1 \int_0^\infty \frac{N_0}{2} \delta(\lambda_2 - \lambda_1 - \tau) h(\lambda_1) h(\lambda_2) d\lambda_2 \\ &= \frac{N_0}{2} \int_0^\infty h(\lambda) h(\lambda + \tau) d\lambda = \frac{N_0}{2} R_h(\tau) \end{aligned}$$

De la anterior expresión, se observa que en presencia del RBG a la entrada de un sistema lineal e invariante en el tiempo, la función de correlación de la señal aleatoria de salida es proporcional a una dependencia, que de manera condicional, se puede denominar función correlación de la respuesta a impulso del sistema, esto es,

$$R_\eta(\tau) = S_\xi(\omega) R_h(\tau)$$

El análisis en el tiempo de la función de correlación de salida se puede realizar integrando hasta el valor  $t$ ,

$$R_\eta(t_1, t_2) = \frac{N_0}{2} \alpha^2 e^{-\alpha(t_1+t_2)} \int_0^{t_1} \int_0^{t_2} e^{\alpha(\lambda_1+\lambda_2)} \delta(\lambda_1 - \lambda_2) d\lambda_2 d\lambda_1$$

Por cuanto, en el desarrollo de la integral doble es necesario tener en cuenta que es cierta la siguiente expresión:

$$\int_{\lambda_0-\varepsilon}^{\lambda_0+\varepsilon} g(\lambda_0 - \lambda) d\lambda = g(\lambda_0)$$

solo para  $\varepsilon > 0$ , cuando el punto específico  $\lambda = \lambda_0$  se encuentra dentro de los límites de integración, entonces, el análisis se lleva a cabo para dos casos:

1.  $t_2 - t_1 > 0$ :

$$\begin{aligned} \int_0^{t_1} \int_0^{t_1} e^{\alpha(\lambda_1+\lambda_2)} \delta(\lambda_1 - \lambda_2) d\lambda_2 d\lambda_1 &= \int_0^{t_1} e^{\alpha\lambda_2} d\lambda_2 \int_0^{t_1} e^{\alpha\lambda_1} \delta(\lambda_1 - \lambda_2) d\lambda_1 \\ &= \int_0^{t_1} e^{2\alpha\lambda_2} d\lambda_2 = \frac{1}{2\alpha} (e^{2\alpha t_1} - 1) \end{aligned}$$

2.  $t_2 - t_1 < 0$ :

$$\int_0^{t_2} \int_0^{t_2} e^{\alpha(\lambda_1+\lambda_2)} \delta(\lambda_1 - \lambda_2) d\lambda_2 d\lambda_1 = \frac{1}{2\alpha} (e^{2\alpha t_2} - 1)$$

Juntando ambas soluciones se obtiene la siguiente función de correlación del proceso de salida:

$$R_\eta(t, \tau) = \frac{\alpha N_0}{4} e^{\alpha|\tau|} (1 - e^{-2\alpha t})$$

La varianza resulta al hacer  $\tau = 0$ :

$$\sigma_{\eta}^2(t) = \frac{\alpha N_0}{4} (1 - e^{-2\alpha t})$$

Como se observa, la varianza tiene la mismo forma de cambio que la media a la salida del ejemplo 5.1; sólo que la varianza crece dos veces más rápido que la media, esto es, tiene un menor intervalo de transición (curva ' $\cdot$ ' para  $\alpha = 2$  de la Figura 5.1).

**Funciones de correlación mutua entrada-salida.** En los sistemas lineales, es de suponer alguna relación de dependencia estadística entre los señales aleatorias de entrada y salida, dada por la respectiva función de correlación mutua, que se considera en dos formas:

$$\begin{aligned} R_{\xi\eta}(\tau) &= E\{\xi(t)\eta(t+\tau)\} = E\left\{\xi(t)\int_0^{\infty}\xi(t+\tau-\lambda)h(\lambda)d\lambda\right\} \\ &= \int_0^{\infty} E\{\xi(t)\xi(t+\tau-\lambda)\}h(\lambda)d\lambda = \int_0^{\infty} R_{\xi}(\tau-\lambda)h(\lambda)d\lambda \end{aligned} \quad (5.9a)$$

$$\begin{aligned} R_{\eta\xi}(\tau) &= E\{\xi(t+\tau)\eta(t)\} = E\left\{\xi(t+\tau)\int_0^{\infty}\xi(t-\lambda)h(\lambda)d\lambda\right\} \\ &= \int_0^{\infty} E\{\xi(t+\tau)\xi(t-\lambda)\}h(\lambda)d\lambda = \int_0^{\infty} R_{\xi}(\tau+\lambda)h(\lambda)d\lambda \end{aligned} \quad (5.9b)$$

Por cuanto la función de convolución dentro de ambas integrales es par con relación a  $\lambda = -\tau$ , y la respuesta a impulso se define solo para valores positivos de  $\lambda$ , entonces, es de esperar que la función de correlación mutua  $R_{\xi\eta}(\tau)$ , en general, se diferencie de  $R_{\eta\xi}(\tau)$ .

**Ejemplo 5.4.** Hallar ambas funciones de correlación mutua entre la entrada y la salida del circuito  $RC$  del ejemplo 1.3, si se tiene una señal de entrada aleatoria  $\xi(t)$  de voltaje, que corresponde a RBG.

Teniendo en cuenta las expresiones (5.9a) y (5.9b), se obtiene

$$\begin{aligned} R_{\xi\eta}(\tau) &= \frac{N_0}{2} \int_0^{\infty} \delta(\tau-\lambda)h(\lambda)d\lambda = \frac{N_0}{2} h(\tau)u(\tau) \\ R_{\eta\xi}(\tau) &= \frac{N_0}{2} \int_0^{\infty} \delta(\tau+\lambda)h(\lambda)d\lambda = \frac{N_0}{2} h(-\tau)u(-\tau) \end{aligned}$$

Los resultados obtenidos sobre la asimetría de las respectivas funciones de correlación conjunta confirman lo obtenido en el numeral §3.1.4.

### 5.1.2. Análisis en la frecuencia

El análisis matemático en el tiempo presenta dificultades de solución en el desarrollo de los momentos de salida, para el caso de procesos aleatorios con FDP diferente a la Gaussiana, debido a la presencia en la mayoría de ellos de las integrales dobles, de las cuales una de ellos es la correlación. Así por ejemplo, la expresión (5.8) entre las funciones de correlación (3.6b) de entrada y salida en un sistema lineal e invariante en el tiempo, en caso de análisis del régimen estacionario y para una señal aleatoria estacionaria, está dada por la relación:

$$R_{\eta}(\tau) = \int_0^{\infty} d\lambda_1 \int_0^{\infty} R_{\xi}(\lambda_2 - \lambda_1 - \tau) h(\lambda_1) h(\lambda_2) d\lambda_2$$

En este sentido, es preferible realizar el análisis de la transformación de cada uno de los momentos y valores de aleatoriedad sobre el dominio de la frecuencia, a fin de reducir la complejidad del desarrollo matemático.

La representación en la frecuencia de los procesos aleatorios está dada por la DEP, la cual a su vez, se relaciona con la función de correlación a través de las transformadas de Wiener-Jinchin (ecs. (3.27) y (3.28)):

$$S_{\eta}(\omega) = \mathcal{F}\{R_{\eta}(\tau)\} = \int_{-\infty}^{\infty} \left\{ \int_0^{\infty} d\lambda_1 \int_0^{\infty} R_{\xi}(\lambda_2 - \lambda_1 - \tau) h(\lambda_1) h(\lambda_2) \right\} e^{-j\omega\tau} d\tau$$

Cambiando el orden de integración, se obtiene:

$$\begin{aligned} S_{\eta}(\omega) &= \int_0^{\infty} d\lambda_1 \int_0^{\infty} h(\lambda_1) h(\lambda_2) d\lambda_2 \int_{-\infty}^{\infty} R_{\xi}(\lambda_2 - \lambda_1 - \tau) e^{-j\omega\tau} d\tau \\ &= \int_0^{\infty} d\lambda_1 \int_0^{\infty} h(\lambda_1) h(\lambda_2) d\lambda_2 S_{\xi}(\omega) e^{-j\omega(\lambda_2 - \lambda_1)} d\lambda_2 \\ &= S_{\xi}(\omega) \int_0^{\infty} h(\lambda_1) e^{j\omega\lambda_1} d\lambda_1 \int_0^{\infty} h(\lambda_2) d\lambda_2 e^{-j\omega\lambda_2} d\lambda_2 \\ S_{\eta}(\omega) &= S_{\xi}(\omega) H(-\omega)H(\omega) = S_{\xi}(\omega) |H(\omega)|^2 \end{aligned} \quad (5.10)$$

La expresión (5.10) es similar a la obtenida para la DEP de señales en (1.40); analogía que puede ser empleada en los métodos de análisis de funciones de transferencia descritas por funciones racionales, cuando a la entrada se tienen procesos estacionarios.

De la expresión (3.25), se puede encontrar la varianza:

$$\sigma_{\eta}^2 = \frac{1}{2\pi} \int_{-\infty}^{\infty} S_{\eta}(\omega) d\omega = \int_{-\infty}^{\infty} S_{\xi}(\omega) H(\omega) H(-\omega) d\omega \quad (5.11)$$

El valor cuadrático medio, teniendo en cuenta la relación dada en (5.3) para el valor

medio, se halla de la siguiente manera

$$m_{2\eta} = \sigma_{\eta}^2 + m_{1\eta}^2 = \int_{-\infty}^{\infty} S_{\xi}(\omega) H(\omega) H(-\omega) d\omega + m_{1\xi}^2 (H(0))^2$$

**Ejemplo 5.5.** Hallar la DEP de salida para el circuito *RC* del ejemplo 1.3, si a la entrada se tiene RBG  $\xi(t)$ .

Teniendo en cuenta la densidad espectral de potencia del ruido blanco (3.33), entonces

$$S_{\eta}(\omega) = \frac{N_0}{2} |H(\omega)|^2 = \frac{N_0}{2} \frac{\alpha}{j\omega + \alpha} \frac{\alpha}{-j\omega + \alpha} = \frac{N_0}{2} \left| \frac{\alpha^2}{\omega^2 + \alpha^2} \right|$$

La varianza, en este caso, se halla integrando la anterior expresión:  $\sigma_{\eta}^2 = \frac{N_0}{2} \frac{\alpha}{2}$ .

**Ejemplo 5.6.** Dado un proceso estacionario  $\xi(t)$ , con valor medio  $m_{1\xi}$  y función de correlación  $R_{\xi} = \sigma_{\xi}^2 r_{\xi}(\tau)$ , a la entrada de un dispositivo de promedio deslizante con apertura finita de tiempo  $(t-T, t+T)$ . Hallar el cambio de los respectivos valores medios.

El proceso a la salida se determina como

$$\eta(t) = (2T)^{-1} \int_{t-T}^{t+T} \xi(\lambda) d\lambda \quad (1)$$

con valor medio

$$m_{1\eta} = E\{\eta(t)\} = \frac{1}{2T} \int_{t-T}^{t+T} E\{\xi(\lambda)\} d\lambda = m_{\xi}$$

y función de correlación

$$R_{\eta}(\tau) = E\{\eta(t) - m_{1\eta}, \eta(t+\tau) - m_{1\eta}\} = \frac{1}{4T^2} \int_{t-T}^{t+T} \int_{t+\tau-T}^{t+\tau+T} R_{\xi}(u-v) du dv,$$

El núcleo de la integral se reemplaza por la expresión:

$$R_{\xi}(u-v) = \frac{1}{2\pi} \int_{-\infty}^{\infty} S_{\xi}(\omega) \exp(j\omega(u-v)) d\omega$$

se obtiene

$$\begin{aligned} R_{\eta}(\tau) &= \frac{1}{2\pi 4T^2} \int_{-\infty}^{\infty} S_{\xi}(\omega) d\omega \int_{t-T}^{t+T} \exp(-j\omega v) dv \int_{t+\tau-T}^{t+\tau+T} \exp(j\omega u) du \\ &= \frac{1}{2\pi} \int_{-\infty}^{\infty} S_{\xi}(\omega) (\text{sinc}(\omega T))^2 d\omega \end{aligned} \quad (2)$$

Por lo tanto, el espectro de salida es de la forma,  $S_\eta(\omega) = S_\xi(\omega) |\text{sinc}(T\omega)|^2$ , que se concentra en el lóbulo principal de la función sinc, en la cercanía de los valores  $\omega T$ . De esta manera, se eliminan las componentes espectrales superiores, esto es, se eliminan los cambios rápidos del proceso  $\xi(t)$  en el tiempo.

La comparación de (5.1) con (1) muestra que la operación de promedio deslizando en el tiempo se analiza como el paso de una señal aleatoria  $\xi(t)$  a través de un sistema lineal con respuesta a impulso y función de frecuencia, dadas respectivamente como,

$$h(t) = \begin{cases} 1/2T, & |t| < T \\ 0, & |t| > T \end{cases}, \quad H(\omega) = \text{sinc}(\Delta\omega)$$

Reemplazando en (2),  $S_\xi(\omega) = \mathcal{F}\{R_\xi(\lambda)\}$  e intercambiando el orden de integración, se obtiene la función de correlación de salida,

$$R_\xi(\tau) = \frac{1}{2T} \int_{-2T}^{2T} \left(1 - \frac{|\lambda|}{2T}\right) R_\xi(\tau - \lambda) d\lambda \quad (3)$$

La varianza del proceso a la salida se halla asumiendo en (3),  $\tau = 0$ ,

$$\sigma_\eta^2 = R_\eta(0) = \frac{1}{2T} \int_{-2T}^{2T} \left(1 - \frac{|\lambda|}{2T}\right) R_\xi(\lambda) d\lambda$$

Entonces para el cálculo de la varianza del promedio deslizando en el tiempo de una señal estacionaria en el sentido amplio, es necesario conocer su función de correlación,

$$\sigma_\eta^2 = R_\eta(0) = \frac{1}{2T} \int_{-2T}^{2T} \left(1 - \frac{|\lambda|}{2T}\right) R_\xi(\lambda) d\lambda = \frac{\sigma_\xi^2}{2T} \int_{-2T}^{2T} \left(1 - \frac{\tau}{T}\right) r_\xi(\tau) d\tau \quad (4)$$

La expresión (4) presenta interés para los modelos de señales aleatorios, en la medida en que se analiza su valor asintótico,  $\lim_{T \rightarrow \infty} \sigma_\eta^2(T) = 0$ .

### 5.1.3. Empleo de operadores lineales

En la sección §3.1.1 se analiza la transformación de variables empleando operadores lineales. Un sistema lineal e invariante en el tiempo descrito por la ecuación diferencial (1.38), se puede representar empleando operadores en la siguiente forma:

$$\mathcal{A}\{\eta(t)\}(p, t) = \mathcal{B}\{\xi(t)\}(p, t) \quad (5.12)$$

donde  $\mathcal{A}(p, t) = \sum_m \alpha_m p^m$  y  $\mathcal{B}(p, t) = \sum_n \beta_n p^n$  son los operadores de salida y entrada respectivamente, siendo  $p = d/dt$ . De la expresión (5.12), se obtiene la relación formal entre la señal de entrada y salida, mediante los respectivos operadores:

$$\eta(t) = \frac{\mathcal{B}(p, t)}{\mathcal{A}(p, t)} \{\xi(t)\} = \mathcal{K}\{\xi(t)\}(p, t)$$

Empleando el operador  $\mathcal{K}(p, t)$  para la descripción del sistema, se pueden hallar las respectivas expresiones para los siguientes momentos estadísticos:

$$m_{1\eta}(t) = \mathcal{K}\{m_{1\xi}(t)\}(p, t) \quad (5.13a)$$

$$R_{\eta}(t_1, t_2) = \mathcal{K}\{\mathcal{K}\{R_{\xi}(t_1, t_2)\}(p, t_2)\}(p, t_1) \quad (5.13b)$$

**Ejemplo 5.7.** Sea un sistema diferenciador de primer orden, a cuya entrada se aplica la señal aleatoria  $\xi(t)$  con valor medio,  $m_{1\xi}(t) = \sin \omega t$ , y función de correlación de la forma  $R_{\xi}(t_1, t_2) = \sigma_{\xi}^2 \exp(-\alpha(t_2 - t_1))^2$ . Determinar la media y la varianza del proceso a la salida.

La señal aleatoria a la salida  $\eta(t)$  se relaciona con la entrada, mediante el operador lineal,  $\eta(t) = \mathcal{K}\{\xi(t)\}(p, t) = d\xi/dt$ , entonces, de las relaciones (5.13a) y (5.13b) se obtiene:

$$m_{1\eta}(t) = \frac{d}{dt} m_{1\xi}(t) = \omega \cos \omega t$$

$$R_{\eta}(t_1, t_2) = \frac{\partial^2}{\partial t_1 \partial t_2} R_{\xi}(t_1, t_2) = 2\sigma_{\xi}^2 \alpha e^{-\alpha(t_2 - t_1)^2} (1 - 2\alpha(t_2 - t_1)^2)$$

Haciendo  $t_2 = t_1 = t$  se encuentra  $R_{\eta}(t, t) = \sigma_{\eta}^2 = 2\alpha\sigma_{\xi}^2$ .

**Ejemplo 5.8.** Sea un sistema integrador,  $\eta(t) = \int_0^t \xi(\lambda) d\lambda$  a la entrada del cual se aplica una señal aleatoria  $\xi(t)$  con valor medio  $m_{1\xi}(t)$  y función de correlación  $R_{\xi}(\tau)$ . Determinar el valor medio y la función de correlación del proceso a la salida.

De manera similar al caso anterior, teniendo en cuenta que la integral corresponde a un operador lineal, se obtiene la relación entre los momentos de entrada y salida,

$$m_{1\eta}(t) = \int_0^t m_{1\xi}(t) dt$$

Haciendo  $m_{1\xi}(t) = m_{1\xi} = \text{const}$ , se observa que la media a la salida resulta variable en el tiempo,  $m_{1\eta}(t) = m_{1\xi}t$  (excepto para el caso  $m_{1\xi} = 0$ ), luego el proceso a la salida presenta un comportamiento no estacionario. A efecto de simplificar el análisis se asume  $m_{1\xi}(t) = 0$ , entonces la función de correlación, a partir de las relación (5.13b) resulta en:

$$R_{\eta}(t_1, t_2) = \int_0^{\lambda_1} \int_0^{\lambda_2} R_{\xi}(t_1, t_2) dt_1 dt_2$$

la cual muestra, que para integrar una señal aleatoria se exige la existencia de su función de correlación. Asumiendo además, la estacionariedad del proceso a la entrada, la función de correlación del proceso a la salida se resume hasta la integral doble

$$R_{\eta}(t_1, t_2) = \int_0^{\lambda_1} \int_0^{\lambda_2} R_{\xi}(t_1 - t_2) dt_1 dt_2$$

que depende por separado de  $t_2$  y  $t_1$  y no de su diferencia, por lo tanto, el proceso a la salida no es estacionario.

La varianza se puede obtener de la última integral, teniendo en cuenta (3.12b) [12]:

$$\sigma_{\eta}^2(t) = 2 \int_0^t (t - \tau) R_{\xi}(\tau) d\tau$$

En general, se puede decir que las FDP cambian a su paso por operadores lineales, excepto para el caso Gaussiano, que mantiene su estructura.

## Problemas

**Problema 5.1.** Se tiene una mezcla lineal,  $x(t) = s(t) + \eta(t)$ , ambos procesos con correlación mutua nula, donde  $s(t) = s_0 \cos(\omega t + \phi)$  es la señal útil y  $\eta(t)$  es el ruido con densidad normal  $\mathcal{N}(m_{1\eta}(t), \sigma_{\eta}(t))$ . Hallar la FDP de la suma.

**Problema 5.2.** Un proceso aleatorio  $\eta(t)$  está compuesto de una cantidad  $N \gg 1$  de señales aleatorias independientes  $\xi_i$ , cada una de ellas con FDP Rayleigh. Hallar la FDP del proceso suma.

**Problema 5.3.** Dados los procesos  $\xi_1(t) = a \cos \omega_c t$  y  $\xi_2(t) = b \sin \omega_c t$ , siendo  $\omega_c = \text{const}$ , y  $a, b$  variables aleatorias invariables en el tiempo. Hallar la esperanza, la función de correlación del proceso suma  $\eta(t) = \xi_1(t) + \xi_2(t)$  y establecer las condiciones para su estacionariedad.

**Problema 5.4.** Sea el proceso aleatorio  $\xi(t)$  con función de correlación  $R_{\xi}(\tau)$ , para el cual se define el proceso diferencial como  $\xi_{\Delta}(t) = \xi(t + \Delta t) - \xi(t)$  que contiene la información de la diferencia de valores del proceso aleatorio  $\xi(t)$  distanciados en el tiempo por el intervalo  $\Delta t$ . Demostrar las siguientes igualdades (donde  $R_{\Delta}(\tau)$  es la función de correlación del proceso diferencial):

$$R_{\Delta}(\tau) = 2R_{\xi}(\tau) - R_{\xi}(\tau + \Delta) - R_{\xi}(\tau - \Delta), \quad S_{\Delta}(\omega) = 4 \sin^2 \frac{\omega \Delta}{2} S_{\xi}(\omega)$$

**Problema 5.5.** Hallar la función de correlación mutua entre el proceso aleatorio  $\xi(t)$  y su derivada  $d\xi/dt$ , además hallar su relación con la respectiva DEP  $S_{\xi}(\omega)$ .

**Problema 5.6.** Sea la función de correlación de un proceso aleatorio  $\xi(t)$  de la forma  $R_{\xi}(\tau) = \sigma_{\xi}^2 \exp(-\alpha^2 \tau^2)$ . Hallar la función de correlación de la derivada del proceso  $d\xi/dt$  y calcular sus valores máximo y mínimo. Además hallar la función de correlación del proceso  $\eta(t) = \xi(t) + d\xi/dt$ .

**Problema 5.7.** Encontrar la diferencia en las funciones de correlación para el caso de RBG y ruido rosado a la salida a la salida del circuito RC (ejemplo 5.1).

**Problema 5.8.** Hallar la función de correlación de la integral del proceso aleatorio  $\xi(t)$  con función de correlación  $R_{\xi}(\tau) = \sigma_{\xi}^2 \exp(-\alpha |\tau|)$ .

**Problema 5.9.** Hallar la función de correlación y la respectiva DEP a la salida de cada uno de los circuitos del problema 1.12, cuando a la entrada se tiene RBG.

**Problema 5.10.** Hallar la función de correlación de salida del circuito mostrado en el ejemplo 5.1, si la función de correlación de entrada es la expresión  $R_{\xi}(\tau) = \frac{\alpha N_0}{2} \exp(-\alpha |\tau|)$

**Problema 5.11.** Sea el RBG a la entrada de un elevador al cuadrado ( $\eta(t) = \xi^2(t)$ ). Encontrar la FDP, la media y la varianza a la salida del sistema no lineal.



## 5.2. Paso de señales aleatorias por sistemas no lineales

Como en el caso de los sistemas lineales, el paso de procesos aleatorios por sistemas no lineales consiste en la determinación de las características de aleatoriedad de los procesos a la entrada y salida del sistema. Sin embargo, debido a la complejidad de solución en este caso, el análisis se enfoca en dos tareas principales: primero, dadas las funciones de no linealidad hallar la representación adecuada de la no linealidad de los sistemas y, segundo, determinar de forma aproximada los momentos y valores de aleatoriedad.

En la primera tarea, por cuanto no existen métodos regulares de solución de cualquier ecuación diferencial no lineal, pues tampoco se tienen métodos regulares de solución de ecuaciones diferenciales no lineales estocásticas. En este sentido, se ha desarrollado una serie de métodos de solución aproximada, entre los cuales están los siguientes:

### 5.2.1. Métodos de linealización

**Linealización determinística.** El método consiste en el cambio de la ecuación diferencial de tipo no lineal que describe el sistema por una (o varias) ecuaciones lineales diferenciales, dado algún criterio de equivalencia de representación. La condición principal del método exige que la no linealidad del sistema sea relativamente pequeña, lo que tiene lugar en el caso de análisis de procesos aleatorios de baja potencia (análisis de señal pequeña), tanto que las fluctuaciones en los incrementos del proceso aleatorio de salida sean menores que el valor estacionario de la reacción  $\eta_0$ , esto es,  $\sigma_\eta \ll \eta_0$ . El valor  $\eta_0$  se supone conocido, y usualmente corresponde al valor asintótico del parámetro en consideración (análisis de estado estable).

La ecuación de estacionariedad, denominada *ecuación de aproximación cero*, se define como la esperanza matemática de la función de no linealidad  $f$ , para un comportamiento medio de entrada

$$\mathcal{K}_0 \{ \eta_0 \} = E \{ f \{ \xi \} \}$$

donde  $\mathcal{K}_0$  es el funcional, que se obtiene del operador  $\mathcal{K}$  para la ecuación diferencial inicial  $\mathcal{K} \{ \eta(t) \} = f \{ \xi(t) \}$  con la condición de que todas sus derivadas sean iguales a cero. La ecuación de incremento diferencial expresada en términos de los incrementos de la salida, dados mediante la diferencia,  $\Delta\eta = \eta - \eta_0$ , se obtiene de la forma

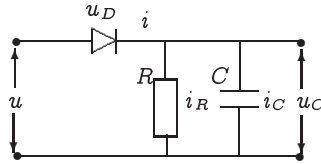
$$\mathcal{K}_0 \{ \eta_0 - \Delta\eta \} - \mathcal{K}_0 \{ \eta_0 \} = f \{ \xi \} - E \{ f \{ \xi \} \} \quad (5.14)$$

Seguidamente, se descompone la función no lineal con argumento  $\Delta\eta$  en la vecindad de  $\eta_0$  por alguna serie de potencia, conservando solo su primer término, de tal manera que se asegure la linealidad de la ecuación diferencial con respecto al argumento  $\Delta\eta$  (*aproximación cero*). La ecuación diferencial de aproximación obtenida sirve para el cálculo de los valores de aleatoriedad de los incrementos de  $\Delta\eta$ , de los cuales una vez obtenidos se calculan los correspondientes valores medios de la salida,  $\eta = \eta_0 + \Delta\eta$ .

**Ejemplo 5.9.** Dado el circuito del detector de amplitud, mostrado en la Figura 5.2, al cual le corresponde la ecuación diferencial

$$\dot{u}_c + \frac{1}{RC}u_c = \frac{1}{C}f(u - u_c),$$

cuyo elemento no lineal, el diodo, se aproxima por la función  $f(\lambda) = i_0 \exp(a\lambda)$   $a = \text{const.}$  Hallar la aproximación en forma de ecuación diferencial lineal, mediante el método de linealización determinística y su solución estacionaria para una entrada aleatoria  $u$  con FDP  $\mathcal{N}(0, \sigma_u)$ , asumiendo que  $\sigma_{u_c} \ll u_{C_0}$ .



**Figura 5.2.** Detector de amplitud

El método de linealización se desarrolla por los siguientes pasos:

1. Desarrollo del modelo a través de la diferencial no lineal. En este caso, la ecuación diferencial no lineal de partida, que tiene la forma

$$\dot{u}_c + \frac{1}{RC}u_c = \frac{1}{C}f(u - u_c) = \frac{1}{C}i_0 \exp(a(u - u_c)), \quad a = \text{const}$$

se puede replantear, de tal manera que en la parte derecha no esté presente la salida buscada,

$$\begin{aligned} \dot{u}_c \exp(au_c) + \frac{1}{RC}u_c \exp(au_c) &= \frac{i_0}{C} \exp(au) \\ \dot{v} + \frac{1}{RC}v \ln v &= \frac{ai_0}{C} e^{au}, \quad v = e^{au_c} \end{aligned} \quad (1)$$

2. Se halla la ecuación promediada de estacionariedad o ecuación de aproximación cero, haciendo en la anterior ecuación  $\dot{v}$  y promediando la parte derecha,

$$\frac{1}{RC}v_0 \ln v_0 = \frac{ai_0}{C} E\{e^{au}\}, \quad (2)$$

De la FDP  $p(u)$ , se halla FDP transformada  $p(v_1)$ , y luego se calcula el valor promedio  $E\{v_1\} = E\{\exp(au)\}$ . Sin embargo, en este caso, debido a la relación exponencial entre las variables, es preferible el empleo de la función característica. Haciendo el valor  $j\omega = a$  se obtiene

$$E\{e^{au}\} = E\{e^{j\omega u}\} = \theta_u(j\omega) = \exp\left(-\frac{\sigma_u^2}{2}\omega\right) = \exp\left(-\frac{\sigma_u^2 a^2}{2}\right)$$

El reemplazo de la última expresión en la ecuación (2) resulta en

$$v_0 \ln v_0 = aRi_0 \exp(\sigma_u^2 a^2 / 2) = au_{C_0} \exp(au_{C_0})$$

por cuanto,  $\exp(au_{C_0}) = v_0$ ,  $au_{C_0} = \ln v_0$ .

La ecuación obtenida se resuelve en términos de  $u_{C_0}$ , y consecuentemente, luego del cálculo de  $\sigma_{u_C}$  se comprueba el cumplimiento de la condición,  $\sigma_{u_C} \ll u_{C_0}$ .

3. Se halla la ecuación no lineal (5.14), combinando (1) con (2),

$$\dot{v} + \frac{1}{RC} v \ln v - \frac{1}{RC} v_0 \ln v_0 = \frac{a i_0}{C} (e^{au} - E\{e^{au}\})$$

de la cual se obtiene la respectiva ecuación para el incremento  $\Delta v = z$ .

Asumiendo que  $v = v_0 + \Delta v = v_0 + z$ , se obtiene

$$RC \dot{z} + (v_0 + z) \ln(v_0 + z) - v_0 \ln v_0 = a R i_0 (e^{au} - E\{e^{au}\}) \quad (3)$$

La ecuación no lineal  $v \ln v = (v_0 + z) \ln(v_0 + z)$  se descompone mediante la serie de Taylor en potencias de  $z$ ,

$$v \ln v = v_0 \ln v_0 + (1 + \ln v_0) z + \frac{1}{2v_0} z^2 - \dots$$

con lo cual, la ecuación de incremento (3) toma la forma,

$$RC \dot{z} + (1 + \ln v_0) z + \frac{1}{2v_0} z^2 - \dots = a R i_0 (e^{au} - E\{e^{au}\})$$

4. Se linealiza la ecuación de incremento, limitando su representación hasta el primer término:

$$\dot{z} + \frac{\ln v_0 + 1}{RC} z = \frac{a}{C} i_0 (e^{au} - E\{e^{au}\})$$

5. Se resuelve el modelo de ecuación diferencial lineal obtenido. Así, por cuanto,

$$\frac{dy}{dx} + Py = Q(x)$$

entonces, la integral general se halla de la expresión

$$y = \exp \left\{ - \int P dx \right\} \left( \int Q(x) \exp \left( \int P dx \right) dx + c \right), \quad P = \frac{\ln v_0 + 1}{RC}$$

Asumiendo la estacionariedad del proceso, entonces, se obtiene el valor asintótico para

$$t \rightarrow \infty, \quad c \exp \left( \int P dt \right) = c \exp(Pt) \rightarrow 0$$

Se debe tener en cuenta que la integración se realiza en el intervalo  $(-\infty, t)$ ,

$$z = e^{-Pt} \int_{-\infty}^t Q(\lambda) e^{P\lambda} d\lambda = \int_{-\infty}^t Q(\lambda) e^{-P(t-\lambda)} d\lambda$$

6. Se hallan las características de aleatoriedad del incremento  $\Delta v$ , de acuerdo a lo expuesto en el numeral §5.1.

**Linealización estadística.** El método, como en el caso anterior de linealización, reemplaza la transformación no lineal por una lineal, dados los criterios de equivalencia probabilística de la solución de ambas ecuaciones.

Sea la señal de salida  $\zeta(t)$  de un elemento con dependencia no lineal invariable en el tiempo,  $\zeta(t) = f(\eta(t))$ , la cual se propone reemplazar mediante la dependencia lineal  $\hat{\zeta} = k_0 + k\eta$ , siendo  $k$  el coeficiente equivalente de ganancia. Escogiendo adecuadamente un criterio de linealización se pueden determinar los coeficientes  $k, k_0$ , así por ejemplo, al tomar el criterio del mínimo error cuadrático medio (2.64) se tiene,

$$\varepsilon^2 = \min_{k_0, k} E \left\{ (\zeta - \hat{\zeta})^2 \right\} = E \left\{ (f(\eta) - k_0 - k\eta)^2 \right\}$$

con lo cual resulta el siguiente sistema de ecuaciones algebraicas,

$$\begin{aligned} \frac{\partial \varepsilon^2}{\partial k^2} &= 2(km_{2\eta} + k_0m_{1\eta} - E\{\zeta\eta\}) = 0, \\ \frac{\partial \varepsilon^2}{\partial k_0^2} &= 2(km_{1\eta} + k_0 - m_{1\zeta}) = 0 \end{aligned}$$

por lo tanto,

$$k = E\{(f(\eta) - m_{1\zeta})(\eta - m_{1\eta})\} / \sigma_\eta^2, \quad k_0 = m_{1\zeta} - km_\eta \quad (5.15)$$

La determinación de los valores de las respectivas constantes  $k_0, k$  en la expresión (5.15) implica el conocimiento de las funciones de los momentos de aleatoriedad del proceso de salida  $\eta(t)$ , lo cual en la práctica es parte de la tarea a calcular y limitando así el empleo de este método. Frecuentemente, una forma de solución aproximada está en asumir la FDP Gaussiana del proceso  $\eta(t)$ , que de manera evidente limita el alcance de la solución.

Una simplificación del método de linealización estadística consiste en el reemplazo de la ecuación diferencial no lineal por una lineal, cambiando los dispositivos de transformación variable en el tiempo por un equivalente invariable.

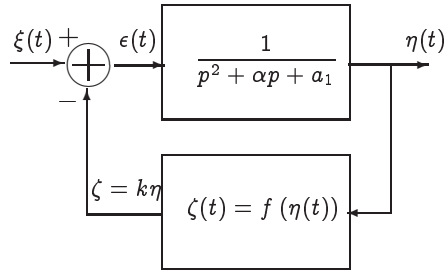
El criterio de equivalencia supone la invariabilidad de los valores de aleatoriedad del proceso en el tiempo, en particular se asume la restricción,

$$\tau_\xi \gg \tau_s \quad (5.16)$$

donde  $\tau_\xi$  es el tiempo de correlación de la señal de entrada y  $\tau_s$  es la constante de tiempo del sistema. Cabe anotar, que la condición (5.16) implica que no necesariamente se debe tener en cuenta el término asociado a  $\dot{\eta}$  con respecto a  $\eta$ .

---

**Ejemplo 5.10.** Sea el sistema no lineal de segundo orden, mostrado en la Figura 5.3, con señal de entrada  $\xi(t)$ , que corresponde al RBG con  $m_{1\xi} = 0$ . Hallar los valores de la media y la varianza del proceso  $\eta(t)$  a la salida.

**Figura 5.3.** Sistema lineal de segundo orden

Por cuanto  $m_{1\xi} = 0$ , entonces,  $m_{1\eta} = m_{1\zeta} = 0$  y la expresión (5.15) se simplifica:

$$k = E\{\eta(t) f(\eta(t))\} / \sigma_\eta^2, \quad k_0 = 0$$

Si se asume la FDP Gaussiana del proceso  $\eta(t)$ ,  $\mathcal{N}(0, \sigma_\eta)$ , además, para efectos de simplicidad, si la no linealidad se restringe a ordenes impares, esto es,  $\zeta = \sum_{i=2}^n a_{2i-1} \eta^{2i-1}$ , entonces la ganancia equivalente es:

$$k = E\{\eta f(\eta)\} / \sigma_\eta^2 = \sum_{i=2}^n a_{2i-1} E\{\eta^{2i}\} / \sigma_\eta^2$$

De otra parte, teniendo en cuenta la expresión para los momentos pares de la FDP normal, entonces,

$$k = 1 \cdot 3 a_3 \sigma_\eta^2 + \dots + 1 \cdot 3 \dots (2n-1) a_{2n-1} (\sigma_\eta^2)^{n-1}$$

El sistema linealizado de la Figura 5.3 tiene función de transferencia,

$$H(\omega) = 1 / \left( (j\omega)^2 + \alpha j\omega + a_1 + k \right)$$

que asumiendo la estacionariedad del proceso  $\eta(t)$ , implica la DEP,  $S_\eta(\omega) = (N_0/2) |H(\omega)|^2$ , de la cual se puede hallar el respectivo valor de varianza,

$$\begin{aligned} \sigma_\eta^2 &= \frac{1}{2\pi} \int_{-\infty}^{\infty} S_\eta(\omega) d\omega = \frac{N_0}{4\pi} \int_{-\infty}^{\infty} |H(\omega)|^2 d\omega \\ &= \frac{N_0}{4\alpha(a_1 + k)} \end{aligned}$$

**Ejemplo 5.11.** Sea el detector de amplitud representado en la Figura 5.2, en cuya entrada se tiene la señal  $u = u_0 \cos(\omega_0 t + \phi(t))$ , que cambia muy lentamente en el tiempo (señal de banda angosta), tal que tiene se puede asumir,  $RC \ll \tau_u$ ,  $RC \gg 2\pi/\omega_0$ , donde  $\tau_u$  es el tiempo de correlación de la envolvente. Hallar el modelo de transformación invariable y equivalente, dado con relación al voltage de salida.

Teniendo en cuenta las siguientes expresiones que describen el sistema,

$$i = i_R + i_C, \quad i_C = C \frac{du_C}{dt}, \quad i_R = \frac{1}{R} u_C, \quad u_D = u - u_C, \quad i = f(u_d)$$

entonces, la ecuación diferencial inicial con términos variables en el tiempo tiene la forma,

$$\begin{aligned} f(u - u_C) &= \frac{1}{R} u_C + C \frac{du_C}{dt} \\ \frac{du_C}{dt} + \frac{1}{RC} u_C &= \frac{1}{C} f(u - u_C) \end{aligned}$$

El modelo de transformación invariable en el tiempo exige analizar el término de la derivada para su eliminación. La condición de un tiempo de correlación de la envolvente de entrada mucho mayor que la inercia del sistema, para el caso de análisis de una señal senoidal con frecuencia  $\Omega$ , permite tomar la siguiente aproximación,

$$\Omega \approx \frac{1}{\tau_u} \quad (1)$$

Sea  $u_C = U_{C_0} \sin \Omega t$ , entonces,  $du_C/dt = \Omega u_{C_0} \cos \Omega t$ . Al tomar la aproximación (1), se tiene

$$\frac{u_{C_0}}{\tau_u} \cos \Omega t + \frac{u_{C_0}}{RC} \sin \Omega t = \frac{1}{C} f(u - u_C)$$

Teniendo en cuenta la condición  $\tau_u \gg RC$ , se observa que el primer término de la ecuación se puede eliminar, por lo tanto, la ecuación diferencial inicial se simplifica hasta el modelo

$$u_c = RC f(u - u_C)$$

que corresponde a una transformación invariable.

Cabe anotar que la otra desigualdad,  $RC \gg 2\pi/\omega_0$ , implica que tiene lugar la filtración de todas las componentes espectrales de alta frecuencia, por lo que se puede asumir la similitud de los anchos de banda de los espectros de las envolventes de entrada y salida.

**Transformación en modelos de Markov.** En aquellos casos en los cuales la salida se puede asumir como un proceso del Markoviano, descrito en el numeral §4.1, la FDP de salida  $p(\eta)$  se puede hallar empleando la ecuación de Fokker-Plank-Kolmogorov (4.34).

La presunción Markoviana se puede asumir si existe la independencia de los incrementos del proceso aleatorio  $\Delta\xi = \xi(t + \tau) - \xi(t)$  para intervalos separados de tiempo. No obstante, esta igualdad no es suficiente. Un criterio cualitativo al respecto puede ser un valor muy pequeño del tiempo de correlación  $\tau_\xi$  de la entrada con respecto a todas las constantes de tiempo del sistema.

Sea un sistema descrito por la más simple ecuación diferencial,  $\dot{\eta} = F(\xi, \eta)$ , dada la entrada  $\eta(t)$  con valor medio 0, entonces, la condición Markoviana (4.20) se asume si tiene lugar la desigualdad

$$\tau_\xi \ll \left( \frac{\partial F}{\partial \xi} \right)^{-1}$$

El análisis de sistemas de mayor orden implica su representación alterna en forma de ecuación vectorial diferencial de primer orden, como se muestra en el caso lineal en (1.61), para el cual se comprueban las respectivas desigualdades.

### 5.2.2. Series de Volterra

Dadas la señal aleatoria de entrada  $\xi(t)$  y salida  $\eta(t)$ , además, asumiendo que de alguna manera se puede determinar los núcleos y, por lo tanto, se puede representar el sistema no lineal con memoria finita por la serie de Volterra (1.44) truncada hasta  $n$  términos,

$$\eta(t) = h_0 + \sum_{k=1}^n \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} h_k(\tau_1, \dots, \tau_k) \xi(t - \tau_1) \cdots \xi(t - \tau_k) d\tau_1 \dots d\tau_k \quad (5.17)$$

entonces, la anterior representación permite analizar el sistema por el método de momentos, por el cual se hallan los momentos de la salida  $\eta(t)$  a partir del conocimiento de los momentos de la señal de entrada  $\xi(t)$ . En este caso, el cálculo de los momentos y valores de aleatoriedad se realiza de manera similar al caso de los sistemas lineales [15].

Así por ejemplo, empleando la operación de promedio a ambos lados de (5.17) se tiene,

$$\overline{\eta(t)} = E\{\eta(t)\} = h_0 + \sum_{k=1}^n \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} h_k(\lambda_1, \dots, \lambda_k) E\{\xi(t - \lambda_1) \cdots \xi(t - \lambda_k)\} d\lambda_1 \dots d\lambda_k$$

Así mismo, para el caso del segundo momento inicial, se obtiene,

$$\begin{aligned} \overline{\eta^2(t)} &= h_0^2 + 2h_0 \overline{\eta(t)} \\ &+ E \left\{ \left( \sum_{k=1}^n \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} h_k(\lambda_1, \dots, \lambda_k) \xi(t - \lambda_1) \cdots \xi(t - \lambda_k) d\lambda_1 \dots d\lambda_k \right)^2 \right\} \end{aligned}$$

De lo anterior, se observa que a medida que crece el orden de aproximación  $n$ , significativamente aumenta la dificultad de cálculo, motivo por el cual, el método de representación con series de Volterra, en la práctica, no se considera para valores  $n \gg 1$ .

**Ejemplo 5.12.** Resolver el ejemplo 5.10 mediante las series de Volterra.

El proceso de salida, formalmente se representa por la convolución  $\eta(t) = h(t) * \epsilon(t)$ , siendo

$$\epsilon(t) = \xi(t) - (a_3 \eta^3(t) + \cdots + a_{2n-1} \eta^{2n-1}(t))$$

mientras,  $h(t)$  es la respuesta a impulso de la parte lineal del sistema con función de transferencia,  $H(\omega) = (p^2 + \alpha p + a_1)^{-1}$ . Por lo tanto, se tiene,

$$\begin{aligned} \eta(t) &= \int_{-\infty}^{\infty} h(\tau) \xi(t - \tau) d\tau - a_3 \int_{-\infty}^{\infty} h(\tau) \eta^3(t - \tau) d\tau - a_1 \int_{-\infty}^{\infty} h(\tau) \eta^5(t - \tau) d\tau - \cdots \\ &\quad - a_{2n-1} \int_{-\infty}^{\infty} h(\tau) \eta^{2n-1}(t - \tau) d\tau \end{aligned}$$

Se debe hallar la representación del proceso a la salida en forma de la suma,

$$\eta(t) = \eta_1(t) + \eta_2(t) + \cdots + \eta_k(t) + \cdots$$

donde  $\eta_k(t)$  es la componente  $k$ , que corresponde al  $k$  núcleo. Al igualar ambas identidades anteriores se obtiene,

$$\begin{aligned} \sum_{k=1} \eta_k(t) &= \int_{-\infty}^{\infty} h(\tau) \xi(t-\tau) d\tau - a_3 \int_{-\infty}^{\infty} h(\tau) \eta^3(t-\tau) d\tau \\ &\quad - a_1 \int_{-\infty}^{\infty} h(\tau) \eta^5(t-\tau) d\tau - \cdots - a_{2n-1} \int_{-\infty}^{\infty} h(\tau) \eta^{2n-1}(t-\tau) d\tau \\ &= \int_{-\infty}^{\infty} h(\tau) \xi(t-\tau) d\tau - a_3 \int_{-\infty}^{\infty} h(\tau) \eta_1^3(t-\tau) d\tau \\ &\quad - 3a_3 \int_{-\infty}^{\infty} h(\tau) \eta_1^2(t-\tau) \eta_2(t-\tau) d\tau - 3a_3 \int_{-\infty}^{\infty} h(\tau) \eta_1(t-\tau) \eta_2^2(t-\tau) d\tau - \cdots \end{aligned}$$

Seguidamente, se pueden igualar los términos de igual potencia de  $\xi(t)$ , con lo cual se obtiene el siguiente sistema de expresiones recurrentes:

$$\begin{aligned} \eta_1(t) - \int_{-\infty}^{\infty} h(\tau) \xi(t-\tau) d\tau &= 0, \quad \eta_2(t) = 0, \\ \eta_3(t) + a_3 \int_{-\infty}^{\infty} h(\tau) \eta_1^3(t-\tau) d\tau &= 0, \quad \eta_4(t) = 0, \\ \eta_5(t) + 3a_3 \int_{-\infty}^{\infty} h(\tau) \eta_1^2(t-\tau) \eta_3(t-\tau) d\tau + a_5 \int_{-\infty}^{\infty} h(\tau) \eta_1^5(t-\tau) d\tau &= 0, \quad \cdots \end{aligned}$$

Del primer término, se puede concluir que  $h_1(\tau) = h(\tau)$ , que al reemplazar en el tercer término resulta en

$$\begin{aligned} \eta_3(t) &= -a_3 \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h(\tau) h(\lambda_1) h(\lambda_2) h(\lambda_3) \xi(t-\tau-\lambda_1) \xi(t-\tau-\lambda_2) \\ &\quad \times \xi(t-\tau-\lambda_3) d\lambda_1 d\lambda_2 d\lambda_3 d\tau = 0, \end{aligned}$$

Se puede demostrar que la función de transferencia de dimensión múltiple del sistema con núcleo  $h_k(\lambda_1, \dots, \lambda_k)$ , que precede a un sistema lineal con función de transferencia  $H(p)$ , se determina por la relación

$$H(p_1 + p_2 + \cdots + p_k) H_k(p_1, p_2, \dots, p_k)$$

Si por el contrario, el sistema lineal se encuentra antes del sistema con núcleo  $h_k(\lambda_1, \dots, \lambda_k)$ , la función de transferencia del todo el sistema será:  $H_k(p_1, p_2, \dots, p_k) H(p_1) H(p_2) \cdots H(p_k)$ .



Teniendo en cuenta la anterior, el tercer término de la descomposición,  $\eta_3(t)$ , se puede analizar como la salida de un sistema, que corresponde a la unión en cascada de un elemento no lineal con núcleo  $h(\lambda_1) h(\lambda_2) h(\lambda_3)$ , y un sistema lineal con respuesta a impulso  $-a_3 h(\tau)$ . Por lo tanto, la función de transferencia de tal sistema será:

$$H_3(p_1, p_2, p_3) = -a_3 H(p_1 + p_2 + p_3) H(p_1) H(p_2) H(p_3)$$

De manera similar, se obtiene la función de transferencia para el quinto término:

$$H_5(p_1, \dots, p_5) = H(p_1 + \dots + p_5) H(p_1) \dots H(p_5) (3a_2^2 H(p_3 + p_4 + p_5) - a_5)$$

A partir de estos primeros términos, se aproxima la correspondiente serie de descomposición para la varianza  $\sigma_\eta^2$  del proceso de salida:

$$\sigma_\eta^2 = E \left\{ \left( \sum_{k=1}^{\infty} \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} h_k(\tau_1, \dots, \tau_k) \xi(t - \lambda_1) \dots \xi(t - \lambda_k) d\lambda_1 \dots d\lambda_k \right) \right. \\ \left. \left( \sum_{k=1}^{\infty} \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} h_k(\tau_1, \dots, \tau_k) \xi(t - \lambda_1) \dots \xi(t - \lambda_k) d\lambda_1 \dots d\lambda_k \right) \right\}$$

Por cuanto el operador de promedio se puede introducir en el integral, entonces,

$$\sigma_\eta^2 = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h_1(\tau) h_1(\lambda_1) E\{\xi(t - \tau) \xi(t - \lambda_1)\} d\tau d\lambda_1 \\ + 2 \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h_1(\tau) h_3(\lambda_1, \lambda_2, \lambda_3) \\ \times E\{\xi(t - \tau) \xi(t - \lambda_1) \xi(t - \lambda_2) \xi(t - \lambda_3)\} d\tau d\lambda_1 d\lambda_2 d\lambda_3 + \dots$$

El proceso a la entrada  $\xi(t)$  es ruido blanco Gaussiano con media 0 y función de correlación  $R_\xi(t_m, t_n) = E\{\xi(t_m) \xi(t_n)\} = (N_0/2) \delta(t_m - t_n)$ , por lo tanto,

$$E\{\xi(t_1) \dots \xi(t_{2n+1})\} = 0, \quad E\{\xi(t_1) \dots \xi(t_{2n})\} = \sum \prod R_\xi(t_j, t_p) R_\xi(t_p, t_m) R_\xi(t_q, t_r)$$

donde la suma se lleva a cabo por todas las particiones de las diferentes parejas, mientras el producto se realiza por todas las parejas de elementos en cada partición.

De esta manera, se tiene que la varianza buscada es igual a,

$$\sigma_\eta^2 = \frac{N_0}{2} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h_1(\tau_1) h_1(\lambda_1) \delta(\lambda_1 - \tau) d\lambda_1 d\tau + \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h_1(\tau) h_3(\lambda_1, \lambda_2, \lambda_3) \\ \times (\delta(\lambda_1 - \tau) \delta(\lambda_2 - \lambda_3) + \delta(\lambda_2 - \tau) \delta(\lambda_3 - \lambda_1) + \delta(\lambda_3 - \tau) \delta(\lambda_1 - \lambda_2)) d\tau d\lambda_1 d\lambda_2 d\lambda_3$$

Sin embargo, el cálculo del anterior integral se simplifica empleando la transformada de Laplace de dimensión múltiple. Así, conocido que,

$$L\{y_i(t_1, \dots, t_i) y_k(t_{i+1}, \dots, t_{i+k})\} = Y_i(p_1, \dots, p_i) Y_k(p_{i+1}, \dots, p_{i+k})$$

donde

$$\begin{aligned} L\{y_k(t_1, \dots, t_k)\} &= Y_k(p_1, \dots, p_k) \\ &= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} y_k(t_1, \dots, t_k) \exp(-p_1 t_1 - \cdots - p_k t_k) dt_1 \dots dt_k \end{aligned}$$

luego,

$$\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} y_i(t_1, \dots, t_i) y_k(t_{i+1}, \dots, t_{i+k}) dt_1 \dots dt_{i+k} = Y_i(0, \dots, 0) Y_k(0, \dots, 0)$$

El cálculo del primer término de la varianza se realiza de la siguiente forma:

$$L\{h_1^2(\lambda)\} = L\{h_1(\lambda_1) h_1(\lambda_2)\}_{p_1 \Leftrightarrow p_2} = \frac{1}{(p_1^2 + \alpha p_1 + a_1)(p_2^2 + \alpha p_2 + a_1)} \Big|_{p_1 \Leftrightarrow p_2}$$

que por tabla se llega a la siguiente expresión:

$$L\{h_1^2(\lambda)\} = \frac{2}{(p_1 + \alpha)(p_1^2 + 2\alpha p_1 + 4a_1)},$$

luego,

$$\int_{-\infty}^{\infty} h_1^2(\lambda) d\lambda = \frac{1}{2\alpha a_1}$$

En el cálculo del segundo término de la varianza, se realiza de forma similar al anterior,

$$L\{h_1(\lambda_1) h_3(\lambda_2, \lambda_3, \lambda_3)\} = -a_3 H(p_1) H(p_2) H(p_2 + p_3) \left( H(p_3) H(p_4) \Big|_{p_3 \Leftrightarrow p_4} \right)$$

Por lo tanto,

$$\int_{-\infty}^{\infty} h_1(\lambda_1) h_3(\lambda_2, \lambda_3, \lambda_3) d\lambda_3 = -\frac{a_3 H(p_1) H^2(p_2)}{2\alpha a_1}$$

De forma similar, se tiene que

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h_1(\lambda_1) h_3(\lambda_1, \lambda_3, \lambda_3) d\lambda_1 d\lambda_3 = -\frac{a_3}{2\alpha a_1} \left( H(p_1) H^2(p_2) \Big|_{p_1 \Leftrightarrow p_2} \right) \Big|_{p_1=0}$$

Al emplear el método de asociación de variables, se obtiene el valor

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h_1(\lambda_1) h_3(\lambda_1, \lambda_3, \lambda_3) d\lambda_1 d\lambda_3 = -\frac{a_3}{2\alpha a_1} \frac{1}{4\alpha a_1^2}$$

Finalmente, como resultado de todos los cálculos, se obtienen los siguientes dos primeros términos de la descomposición en serie de la varianza del proceso a la salida,

$$\sigma_\eta^2 = N_0 / 4\alpha a_1 - 3(a_3/a_1)(N_0/4\alpha a_1)^2 + \dots$$

que en forma general, converge con la expresión obtenida en el ejemplo 5.10.

## Problemas

**Problema 5.12.** Sea una señal de voltaje aleatoria con FDP  $\mathcal{N}(0, \sigma_v)$ , que alimenta el elemento no lineal con función de paso,

$$i(v) = \begin{cases} i_0 - 2vi_0/v_c + v^2i_0/v_c^2, & v \geq v_c \\ 0, & v < v_c \end{cases}$$

Asumiendo que la potencia del proceso es suficientemente pequeña, tal que se cumple la relación,  $|v_c| > 3\sigma_v$ . Hallar la FDP de la corriente a la salida.

**Problema 5.13.** La señal de voltaje aleatoria con FDP  $\mathcal{N}(0, \sigma_v)$  alimenta el elemento no lineal,  $i(v) = Svu(v)$ . Hallar el valor medio y la varianza de la corriente.

**Problema 5.14.** Hallar la función de correlación  $R_i(\tau)$  de la señal de corriente que pasa por un elemento lineal del tipo,  $i(v) = Svu(v)$ , cuando a la entrada del sistema se tiene un proceso estacionario  $\mathcal{N}(0, \sigma_v)$  y función de correlación  $r(\tau)$ .

**Problema 5.15.** Hallar la función de correlación  $R_i(\tau)$ , el valor medio y la varianza de una señal de corriente a la salida de un elemento con característica de paso aproximada por el polinomio  $i = a_0 + a_1v + a_2v^2$ , cuando a la entrada se tiene un proceso estacionario normal  $\mathcal{N}(0, \sigma_v)$ .

**Problema 5.16.** Hallar la función de correlación  $R_y(\tau)$ , del proceso a la salida del elemento con característica

$$y(x) = \begin{cases} -a, & x < 0 \\ a, & x \geq 0 \end{cases}$$

Asumir que  $R_x(0) = 1$ .



### 5.3. Transformación de procesos de Markov

#### 5.3.1. Transformación de señales aleatorias reales por procesos de Markov

En la práctica, el análisis de procesos aleatorios mediante los modelos de Markov exige que se constante la correspondiente condición markoviana, así como la determinación de los coeficientes respectivos de las ecuaciones de Kolmogorov. Aunque en forma estricta, los procesos reales no son del tipo Markoviano para una dimensión, en la medida en que los procesos de Markov no deben cumplir la condición de diferenciabilidad, mientras los reales si presentan esta propiedad, derivada de la existencia obligatoria de inercia en cualquier sistema dinámico físicamente implementable.

En este sentido, para generar los modelos de aproximación de señales reales empleado procesos de Markov con dimensión múltiple, se considera suficiente asumir la estacionariedad de la señal aleatoria, que tenga FDP Gaussiana, con DEP  $S_x(\omega)$  descrita por una función racional quebrada,

$$S_x(\omega) = \frac{|P_m(j\omega)|^2}{|Q_n(j\omega)|^2}, \quad m < n, S_x(\omega) \in \mathbb{Q} \quad (5.18)$$

donde  $Q_n(x) = x^n + \alpha_1 x^{n-1} + \dots + \alpha_n$  y  $P_m(x) = \beta_0 x^m + \beta_1 x^{m-1} + \dots + \beta_m$ , que corresponden al modelo de solución estacionario de la ecuación diferencial

$$\frac{d^n x}{dt^n} + \alpha_1 \frac{d^{n-1} x}{dt^{n-1}} + \dots + \alpha_n x(t) = \beta_0 \frac{d^m \xi}{dt^m} + \dots + \beta_m \xi(t) \quad (5.19)$$

donde  $\xi(t)$  es RBG con función de correlación,  $R_\xi(\tau) = \delta(\tau)$ .

En general, una ecuación diferencial de orden mayor se puede representar en forma de ecuación diferencial de primer orden, como se explica en (1.61), así,

$$\begin{aligned} \frac{dx_n}{dt} + \alpha_n x_1 + \alpha_{n-1} x_2 + \dots + \alpha_1 x_n + c_{n-m} \xi^{(m)}(t) + (c_{n-m+1} + \alpha_1 c_{n-m}) \xi^{(m-1)}(t) + \dots \\ \dots + (c_{n-1} + \alpha_1 c_{n-2} + \dots + \alpha_{m-1} c_{n-m}) \xi^{(1)}(t) + (\alpha_1 c_{n-1} + \alpha_2 c_{n-2} + \dots + \alpha_m c_{n-m}) \xi(t) \\ = \beta_0 \xi^{(m)}(t) + \beta_1 \xi^{(m-1)}(t) + \dots + \beta_{m-1} \xi'(t) + \beta_m \xi(t) \end{aligned} \quad (5.20)$$

donde las constantes  $c_{n-m}, c_{n-m+1}, \dots, c_{n-1}$  se escogen, de tal forma, que al eliminar en (5.19) las derivadas de mayor orden de la función  $x(t)$ , mediante el reemplazo por  $x_k(t)$ , se obtenga una ecuación diferencial de primer orden pero sin términos que contengan derivadas de  $\xi(t)$ .

Al relacionar los coeficientes con igual valor de derivada de  $\xi(t)$  a ambos lados de la ecuación, se obtiene la siguiente expresión recurrente para el cálculo de los coeficientes  $c_{kj}$ :

$$c_k = \beta_{k+m-n} - \sum_{j=1}^{k+m-n} a_j c_{k-j}, \quad k = n-m, n-m+1, \dots, n$$

con lo cual, la ecuación (5.20) se puede escribir en la forma

$$\frac{dx_n}{dt} = - \sum_{j=1}^n \alpha_{n+1-j} x_j + c_n \xi(t) \quad (5.21)$$

El sistema (5.21) contiene  $n$  ecuaciones de primer orden con las correspondientes  $n$  funciones desconocidas  $x, \dots, x_n$ . Además, la función  $\xi(t)$  se mantiene en la parte derecha de la igualdad. Debido a que el comportamiento del conjunto de funciones  $\{x_i\}$  se determina unívocamente por las condiciones iniciales de las mismas, entonces, estas son componentes de un proceso de Markov con dimensión  $n$ . Así mismo, la función inicial  $x(t)$ , por ser  $x(t) = x_i(t)$ , es una componente del proceso de Markov.

Los coeficientes del modelo de Markov,  $a_l$  y  $b_{lm}$ , se obtienen del sistema de ecuaciones (5.21) que tiene la misma estructura de representación de (4.25a), en la cual está presente una sola función  $\xi(t)$ , entonces se obtienen los coeficientes,

$$a_l = \begin{cases} y_{l+1}, & 1 \leq l \leq n-1 \\ - \sum_{j=1}^n \alpha_{n+1-j} y_j, & l = n \end{cases}$$

$$b_{lm} = \begin{cases} 0, & 1 \leq l \leq n-m-1 \\ c_l c_m, & n-m \leq l \leq n \end{cases}, (l \leq m)$$

La última expresión determina  $b_{lm}$ , para cualquiera que sean los índices, debido a la simetría de estos coeficientes.

La transformación a procesos de Markov, además del caso de un proceso normal estacionario con espectro racional quebrado, también es posible extenderlo al análisis de sistemas dinámicos, que contengan elementos (tanto lineales, como no lineales), a la entrada de los cuales estén presentes procesos normales con DEP fraccionales quebradas.

Sea el sistema dinámico descrito por la ecuación:

$$\frac{d^r z}{dt^r} + \gamma_1 \frac{d^{r-1} z}{dt^{r-1}} + \dots + \gamma_r z = f(x, t) \quad (5.22)$$

donde  $r_j$  son las constantes del sistema,  $f(x)$  es la función que describe un elemento no lineal invariante,  $x(t)$  es la señal aleatoria, de carácter estacionario y con estructura Gaussiana, que tiene densidad espectral de potencia en la forma (5.18). Reemplazando el proceso aleatorio  $x(t)$  por uno de Markov con dimensión  $n$ , de acuerdo a la expresión (1.61), y notando cada término como

$$x_{n+1}(t) = z(t), \quad x_{n+2}(t) = \dot{x}_{n+1}, \quad \dots, \quad x_{n+r} = \dot{x}_{n+r-1}(t) \quad (5.23)$$

entonces, la descripción del sistema (5.22) toma la forma:

$$\frac{dx_{n+r}}{dt} = -\gamma_1 x_{n+r} - \gamma_2 x_{n+r-1} - \dots - \gamma_r x_{n+1} + f(x_n, t) \quad (5.24)$$

Las expresiones (5.21), (5.23) y (5.24) conforman un sistema de  $(n + r)$  ecuaciones de primer orden, en la parte derecha de las cuales, además de la función  $x_j$ , está presente solo la señal aleatoria  $\xi(t)$  que tiene las propiedades del ruido blanco Gaussiano (en el sentido angosto). El sistema obtenido de ecuaciones corresponde a un proceso de Markov con dimensión  $n + r$ . En general, la transformación de señales reales a modelos de Markov es viable en los casos, en los que el proceso en análisis se describe por una ecuación diferencial, que puede ser no lineal, que contiene a la entrada una señal aleatoria,  $\xi(t)$ , con propiedades de RBG y que acepta su descomposición en potencias.

---

**Ejemplo 5.13.** Reemplazar el proceso estacionario normal  $x_1(t)$  con valor medio 0 y función de correlación  $R_{x_1}(\tau) = \sigma_x^2 e^{-\alpha|\tau|} \cos \beta \tau$ , por un proceso de Markov de 2 dimensiones.

La respectiva densidad espectral de potencia  $S_x(\omega)$  se determina como:

$$\begin{aligned} S_x(\omega) &= \frac{1}{2} \frac{\sigma^2}{2\pi} \left( \int_{-\infty}^{\infty} e^{-j\omega\tau - \alpha|\tau| + i\beta\tau} d\tau + \int_{-\infty}^{\infty} e^{-j\omega\tau - \alpha|\tau| + i\beta\tau} d\tau \right) \\ &= \frac{1}{2} \frac{\sigma^2}{2\pi} \left( \frac{\alpha}{(\omega - \beta)^2 + \alpha^2} + \frac{\alpha}{(\omega + \beta)^2 + \alpha^2} \right) = \frac{\sigma^2}{\pi} \frac{\omega^2 + \alpha^2 + \beta^2}{(\omega^2 - \beta^2 - \alpha^2)^2 + 4\alpha^2\omega^2} \\ &= \frac{\sigma^2}{\pi} \frac{|j\omega + \sqrt{\alpha^2 + \beta^2}|^2}{|-\omega^2 + 2j\alpha\omega + \alpha^2 + \beta^2|^2} \end{aligned}$$

Por lo tanto,  $x_1(t)$  es la solución estacionaria de la ecuación diferencial:

$$\frac{d^2 x_1}{dt^2} + 2\alpha \frac{dx_1}{dt} + (\alpha^2 + \beta^2)x_1 = \sqrt{2\alpha} \left( \frac{d\xi}{dt} + \sqrt{\alpha^2 + \beta^2}\xi \right) \sigma$$

donde  $\xi(t)$  es ruido blanco Gaussiano.

Cabe anotar que la señal aleatoria  $x_1(t)$  no es del tipo Markoviano, en la medida en que se determina por una ecuación de segundo orden, con lo cual el conocimiento de sus valores iniciales no es suficiente para la determinación de sus propiedades en el tiempo.

Sin embargo, en correspondencia con (1.61) haciendo  $\dot{x}_1(t) = x_2(t) + c_1\xi(t)$ , y realizando el respectivo reemplazo se obtiene:

$$\begin{cases} \frac{dx_2}{dt^2} + c_1 \frac{d\xi}{dt} + 2\alpha x_2 + 2\alpha c_1 \xi + (\alpha^2 + \beta^2)x_1 = \sqrt{2\alpha} \left( \frac{d\xi}{dt} + \sqrt{\alpha^2 + \beta^2}\xi \right) \sigma \\ \frac{d^2 x_1}{dt^2} + 2\alpha \frac{dx_1}{dt} + (\alpha^2 + \beta^2)x_1 = \sqrt{2\alpha} \left( \frac{d\xi}{dt} + \sqrt{\alpha^2 + \beta^2}\xi \right) \sigma \end{cases}$$

si se hace  $c_1 = \sigma\sqrt{2\alpha}$ , en la última expresión la derivada  $d\xi/dt$  desaparece y se obtiene un sistema de dos ecuaciones con dos variables desconocidas, que contienen RBG, y por lo tanto, que determinan el siguiente proceso de Markov con componentes  $x_1(t)$  y  $x_2(t)$ :

$$\begin{cases} \frac{dx_1(t)}{dt} - x_2(t) &= \sigma\sqrt{2\alpha}\xi(t) \\ \frac{dx_2(t)}{dt} + (\alpha^2 + \beta^2)x_1(t) + 2\alpha x_2(t) &= \sigma\sqrt{2\alpha} \left( \sqrt{\alpha^2 + \beta^2} - 2\alpha \right) \xi(t) \end{cases}$$


---

### 5.3.2. Modelos de transformación de señales

En los sistemas reales, las señales de información se pueden aproximar por un proceso aleatorio  $\mathbf{x}(t)$ , con DEP descrita por una función fraccional, (5.18):

$$S_x(\omega) = \frac{N_x}{2} \left| \frac{b_m(j\omega)^m + b_{m-1}(j\omega)^{m-1} + \cdots + b_0}{(j\omega)^n + a_{n-1}(j\omega)^{n-1} + \cdots + a_0} \right|^2 \quad (5.25)$$

en la cual, la condición de realización física del proceso  $\mathbf{x}(t)$  exige que  $m < n$ , esto es, que el proceso tenga potencia finita, además, las raíces del polinomio del denominador deben estar localizadas en la parte negativa real. De lo anteriormente expuesto, un proceso  $\mathbf{x}(t)$  se puede formar al pasar RBG  $\xi(t)$  con valor medio,  $m_{1\xi} = 0$ , y función de correlación  $R_\xi = \frac{N_\xi}{2}\delta(\tau)$ , a través de un filtro lineal con función de transferencia,

$$H_x(p) = \frac{b_m p^m + b_{m-1} p^{m-1} + \cdots + b_0}{p^n + a_{n-1} p^{n-1} + a_{n-2} p^{n-2} + \cdots + a_0} \quad (5.26)$$

donde  $p = d/dt$  es el operador de diferenciación.

En este caso, el proceso Gaussiano estacionario  $\mathbf{x}(t)$  con DEP (5.25), que se obtiene a la salida del filtro (5.26), se describe por la ecuación diferencial de  $n$  orden:

$$\frac{d^n x}{dt^n} + a_{n-1} \frac{d^{n-1} x}{dt^{n-1}} + \cdots + a_0 x(t) = b_m \frac{d^m \xi}{dt^m} + \cdots + b_0 \xi(t), \quad t \geq t_0 \quad (5.27)$$

La ecuación diferencial escalar (5.27) de orden  $n$  se puede reducir a la forma equivalente de una ecuación diferencial vectorial de primer orden,

$$\frac{d\mathbf{x}}{dt} = \mathbf{A}\mathbf{x}(t) + \mathbf{B}\xi(t), \quad \mathbf{x}(t_0) = \mathbf{x}_0, \quad t \geq t_0 \quad (5.28)$$

donde

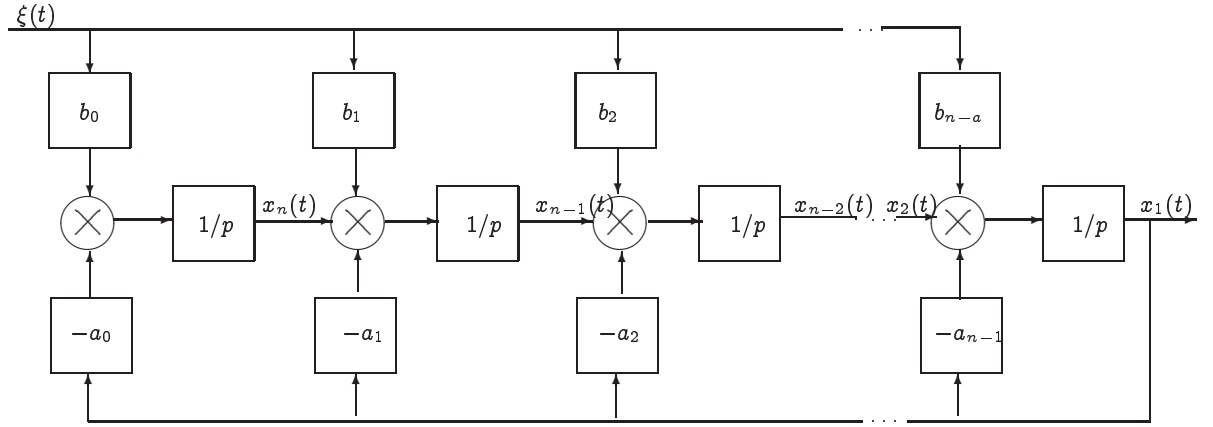
$$\mathbf{A} = \begin{bmatrix} -a_{n-1} & 1 & 0 & 0 & \cdots & 0 \\ -a_{n-2} & 0 & 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & & \vdots \\ -a_0 & 0 & 0 & 0 & \cdots & 0 \end{bmatrix}, \quad \mathbf{x}(t) = \begin{bmatrix} x_1(t) \\ x_2(t) \\ \vdots \\ x_n(t) \end{bmatrix}, \quad \mathbf{B} = \begin{bmatrix} b_m \\ b_{m-1} \\ \vdots \\ b_0 \end{bmatrix} \quad (5.29)$$

siendo  $\mathbf{x}_0$  el vector de condiciones iniciales, que corresponde a los valores del vector de estados  $\mathbf{x}(t)$  en el momento de tiempo  $t_0$ .

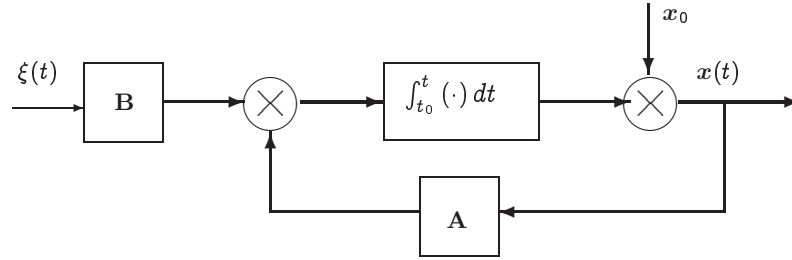
El modelo del filtro con respuesta de frecuencia (5.26) se muestra en la Figura 5.4(a), mientras en la Figura 5.4(b) se muestra el diagrama general del sistema lineal, descrito por la ecuación (5.28), en el cual los vectores se notan con una línea ancha. En forma general, las matrices  $\mathbf{A}(t)$  y  $\mathbf{B}(t)$ , tienen elementos que son funciones del tiempo y se determinan de los parámetros físicos del sistema dinámico.

La ecuación (5.28) describe un proceso Gaussiano de Markov, cuya realización corresponde a la función vectorial  $\mathbf{x}(t)$ , por lo tanto, cualquier proceso, sea estacionario o no,





(a) Filtro formador de proceso de Markov



(b) Diagrama general del sistema de formación

**Figura 5.4.** Modelo formación de procesos de Markov sobre sistemas lineales

pero con DEP fraccional y varianza finita, se puede realizar al pasar RBG por un sistema lineal descrito por la función de transferencia adecuada.

En general, las condiciones iniciales  $\mathbf{x}(t_0)$  en (5.28) son valores aleatorios, con FDP  $p(\mathbf{x}_0)$  estadísticamente independientes de la señal aleatoria  $\xi(t)$ . En principio, el conocimiento de  $p(\mathbf{x}_0)$  es suficiente para la descripción de la ecuación diferencial vectorial (5.28), aunque, también es suficiente la descripción del proceso aleatorio en el intervalo de tiempo  $(t_0, t)$  por la FDP con dimensión  $m$ , tal que  $m \rightarrow \infty$ , esto es, por el funcional de distribución,

$$F\{\mathbf{x}(t)\} = \lim_{\substack{T \rightarrow 0 \\ m \rightarrow \infty}} p_{t_0, t_1, \dots, t_n}(\mathbf{x}_0, \mathbf{x}_1, \dots, \mathbf{x}_m) \quad (5.30)$$

donde  $T = t_{t+\tau} - t_i$ ,  $i = 0, 1, \dots, m$ .

El funcional de distribución (5.30) depende de cada observación del vector aleatorio  $\mathbf{x}(t)$  en la medida en que la FDP  $p_{t_0, t_1, \dots, t_n}(\mathbf{x}_0, \mathbf{x}_1, \dots, \mathbf{x}_m)$  depende de los valores, que toman las variables aleatorias  $\mathbf{x}_0, \dots, \mathbf{x}_m$ .

Los funcionales de FDP existen para una clase limitada de procesos aleatorios, en particular, los Gaussianos y los de Markov. En el caso de los procesos de Markov, la FDP  $p_{t_0, \dots, t_m}(\mathbf{x}_0, \dots, \mathbf{x}_m)$  se puede expresar mediante la densidad de probabilidad inicial  $p_{t_0}(\mathbf{x}_0)$

y la FDP de transición del estado  $\mathbf{x}_{i-1}$  en el momento  $t_{i-1}$  al estado  $\mathbf{x}_i$  en el momento  $t_i$ ,

$$p_{t_i, t_{i+1}}(\mathbf{x}_i | \mathbf{x}_{i-1}) = \pi_{t_i, t_{i-1}}(\mathbf{x}_i | \mathbf{x}_{i-1}), \quad t_{i-1} < t_i, \quad i = 0, 1, \dots, m$$

esto es,

$$p_{t_0, t_1, \dots, t_m}(\mathbf{x}_0, \mathbf{x}_1, \dots, \mathbf{x}_m) = p_{t_0}(\mathbf{x}_0) \prod_{i=1}^m \pi_{t_{i-1}, t_i}(\mathbf{x}_{i-1}, \mathbf{x}_i) \quad (5.31)$$

De (5.31), la FDP con dimensión  $m+1$  para un proceso de Markov, toma la forma,

$$p_{t_0, t_1, \dots, t_m}(\mathbf{x}_0, \mathbf{x}_1, \dots, \mathbf{x}_m) = p_{t_0}(\mathbf{x}_0) \prod_{i=1}^{m+1} \pi_{t_{i-1}, t_i}(\mathbf{x}_{i-1}, \mathbf{x}_i) = \\ \pi_{t_m, t_{m+1}}(\mathbf{x}_{i-1}, \mathbf{x}_i) p_{t_0, t_1, \dots, t_m}(\mathbf{x}_0, \mathbf{x}_1, \dots, \mathbf{x}_m)$$

De lo anterior, se concluye que la FDP inicial más la FDP de transición  $\pi_{t_{i-1}, t_i}(\mathbf{x}_{i-1}, \mathbf{x}_i)$  determinan completamente el proceso vectorial aleatorio de Markov. Por cierto, ambas FDP cumplen, tanto las formas directa e inversa de la ecuación de Kolmogorov.

### 5.3.3. Ejemplos de procesos generados de Markov

En la práctica, se han generado modelos concretos que corresponden a señales reales, las cuales se analizan como procesos de Markov, descritos por (5.28) y por las respectivas FDP  $p_{t_0}(\mathbf{x}_0)$  y  $\pi_{t_{i-1}, t_i}(\mathbf{x}_{i-1}, \mathbf{x}_i)$ . Entre los modelos más conocidos, están los siguientes:

*Señal pasabaja de información.* Sea  $n = 1$ ,  $b_0 = a_0 = \alpha$ . Entonces de (5.28) se obtiene:

$$\frac{dx}{dt} = -\alpha x(t) + \alpha \xi(t), \quad x(t_0) = x_0 \quad (5.32)$$

El modelo (5.32) tiene la misma estructura del circuito  $RC$  presentado en el ejemplo 1.3, con constante de tiempo  $1/\alpha = RC$ , cuando a la entrada se tiene RBG,  $\xi(t)$ . En caso de no existir el primer término de (5.32), se obtiene un proceso estacionario de Wiener.

*Modelo con dos componentes de procesos de Markov de baja frecuencia.*

Sea  $n = 2$ . Entonces, se obtiene el sistema:

$$\begin{aligned} \frac{dx_1}{dt} &= x_2(t) - a_1 x_1(t) + b_1 \xi(t) \\ \frac{dx_2}{dt} &= -a_0 x_1(t) + b_0 \xi(t) \end{aligned} \quad (5.33)$$

En la práctica, es frecuente el empleo de modelos que incluyen dos componentes de procesos de Markov, que tienen una descripción diferente a (5.33), pero que pueden ser descritos por la ecuación diferencial de segundo orden,

$$\frac{d^2 x_1}{dt^2} = a_1 \frac{dx_1}{dt} + a_0 x_1(t) = b_0 \xi(t) + b_1 \frac{d\xi}{dt}$$

que se representa en forma de un sistema de dos ecuaciones lineales diferenciales:

$$\begin{aligned}\frac{dx_1}{dt} &= -\alpha_1 x_1(t) + \alpha_1^2 x_2(t) + \beta_1 \xi(t) \\ \frac{dx_2}{dt} &= -\alpha_2 x_2(t) + \beta_2 \xi(t)\end{aligned}\quad (5.34)$$

donde

$$\alpha_1 + \alpha_2 = a_1, \alpha_1 \alpha_2 = a_0, \alpha_2 \beta_1 = \alpha_1^2, \beta_2 = b_0, b_1 = b_1$$

En particular, cuando  $\alpha_1^2 = \beta_1 = \alpha_2 = \beta_2$ , el proceso aleatorio  $x_1(t)$ , se puede emplear en el modelado de la señal de voz [41]. En este caso,  $x_1(t)$ , se analiza como la salida de un par de circuitos  $RC$  conectados en cascada (sin tener en cuenta su reacción mutua), uno de ellos de baja frecuencia con constante de tiempo  $1/a_2 = R_1 C_1$ , y el otro de alta frecuencia, con  $1/\alpha_1' = R_2 C_2$ , cuando a la entrada de la conexión se tiene RBG. De otra parte, cuando se asumen los valores  $\alpha_1^1 = -\alpha_0'$ ,  $\beta_1 = 0$ ,  $\alpha_2 = \beta_2$ , de (5.34) se obtiene el modelo de trayectorias simples de seguimiento, en el cual el RBG pasa a través de dos circuitos de integración  $RC$  conectados en cascada, con las constantes de tiempo  $1/\alpha_1 = R_1 C_1$  y  $1/\alpha_2 = R_2 C_2$ , respectivamente. Cabe anotar que la conexión en cascada de circuitos  $RC$  se puede generalizar hasta cualquier orden  $n$ , en la medida en que el proceso  $x_1(t)$  se puede representar en la forma (5.34), para cualquiera que sea el orden de la ecuación diferencial (5.27) que lo describe:

$$\frac{dx}{dt} = \mathbf{A}x(t) + \mathbf{B}\xi(t), \quad x(t_0) = x_0, \quad t \geq t_0 \quad (5.35)$$

donde

$$\mathbf{A} = \begin{bmatrix} -\alpha_1 & -\alpha_1^2 & 0 & 0 & \cdots & 0 \\ 0 & -\alpha_2 & -\alpha_2^3 & 0 & \cdots & 0 \\ 0 & 0 & -\alpha_3 & -\alpha_3^4 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & 0 & \cdots & \alpha_n \end{bmatrix}, \quad \mathbf{B} = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_n \end{bmatrix}, \quad x(t) = \begin{bmatrix} x_1(t) \\ x_2(t) \\ \vdots \\ x_n(t) \end{bmatrix}$$

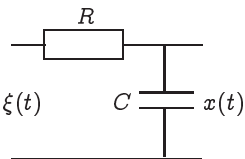
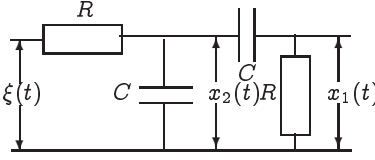
Los coeficientes en (5.27) se relacionan con los valores de (5.35) mediante las siguientes expresiones:

$$\begin{aligned}a_{n-1} &= \sum_{k=1}^n a_k; a_{n-2} = \sum_{i=1}^{n-1} a_i \sum_{j=i+1}^n a_j \\ a_{n-3} &= \sum_{k=1}^{n-2} a_k \sum_{i=k+1}^{n-1} a_i \sum_{j=i+1}^n a_j; a_{n-4} = \sum_{m=1}^{n-3} a_m \sum_{k=m+1}^{n-2} a_k \sum_{i=k+1}^{n-1} a_i \sum_{j=i+1}^n a_j; \\ a_0 &= a_1 a_2 a_3 \cdots a_{n-2} a_{n-1} a_n;\end{aligned}$$

De manera similar, se obtienen los coeficientes para la matriz  $\mathbf{B}$ ,

$$\begin{aligned}
 b_{n-1} &= \beta_1; b_{n-2} = \beta_1 \sum_{j=2}^n a_j - b_2 a_1^{(2)}; \\
 b_{n-3} &= \beta_1 \sum_{i=2}^{n-1} a_i \sum_{j=i+1}^n a_j - \beta_2 \alpha_1^{(2)} \sum_{k=3}^n a_k - \beta_3 a_1^{(2)} a_2^{(3)}; \\
 b_{n-r} &= \beta_1 \sum_{l=2}^{n-r+2} \alpha_l \sum_{s=l+1}^{n-r+3} \alpha_s \dots \sum_{m=u+1}^{n-3} \alpha_m \sum_{k=m+1}^{n-2} \alpha_k \sum_{i=k+1}^{n-1} \alpha_i \sum_{j=i+1}^n \alpha_j - \beta_2 \alpha_1^{(2)} \sum_{s=3}^{n-r+3} \alpha_s \dots \\
 &\quad \sum_{m=u+1}^{n-3} \alpha_m \dots \sum_{j=i-1}^n \alpha_j + \beta_3 \alpha_1^{(2)} \alpha_2^{(3)} \sum_{p=4}^{n-r+4} \alpha_p \dots \sum_{m=n+1}^{n-3} \alpha_m \sum_{k=m+1}^{n-2} a_k \\
 &\quad \sum_{i=k+1}^{n-1} \alpha_i \sum_{j=i+1}^n \alpha_j - \dots - \beta_{r-1} \alpha_1^{(0)} \alpha_2^{(3)} \dots \alpha_{r-2}^{(r-1)} \sum_{i=r+1}^n \alpha_i + \beta_r \alpha_1^{(2)} \alpha_2^{(3)} \dots \alpha_{r-1}^{(r)}
 \end{aligned}$$

El diagrama del filtro de formación, construido en correspondencia con (5.35) se muestra en la Tabla 5.5. Considerando  $B^T = \{0, 0, \dots, \beta_n\}$ ,  $\alpha_i = -\alpha_i^{(i+1)}$ , el filtro obtenido es la conexión en cascada de  $n$  circuitos  $RC$  con constante de tiempo  $1/\alpha_i = R_i C_i$ . Cuando la cantidad de circuitos,  $n$ , es suficientemente grande el filtro se aproxima al tipo Gaussiano.

Modelo del proceso	Función de correlación	Filtro formante
$\frac{dx}{dt} = -\alpha x + \alpha \xi(t),$ $x(t_0) = x_0$	$R_x(\tau) = \sigma_x^2 e^{-\alpha \tau }$	 <p style="text-align: center;"><math>\alpha = 1/RC</math></p>
$\frac{dx_1}{dt} = -\alpha x_1(t) - \alpha x_2(t) + \xi(t),$ $\frac{dx_2}{dt} = -\alpha x_2(t) + \xi(t)$ $x_i(t_0) = x_{i0}, i = 1, 2$	$R_{x_1}(\tau) = \frac{N_x}{8\alpha} (1 - \alpha \tau ) e^{-\alpha \tau }$	 <p style="text-align: center;"><math>\alpha = 1/RC</math></p>

**Figura 5.5.** Modelos de transformación a procesos de Markov

Aunque se tenga un proceso aleatorio real escalar, su representación en forma de un proceso Gaussiano de Markov se realiza en la forma vectorial (5.28), la cual se puede generalizar al caso de análisis conjunto de vectores funcionales continuos con dimensión  $r$ . Cada componente vectorial, a su vez, se modela por el respectivo vector de procesos Gaussianos de Markov  $\mathbf{x}_i(t)$ , descrito por (5.28) en la forma,

$$\frac{d\mathbf{x}_i}{dt} = \mathbf{A}_i(t) \mathbf{x}_i(t) + \mathbf{B}_i(t) \xi_i(t), \quad \mathbf{x}_i(t_0) = \mathbf{x}_{i0}$$

donde  $\mathbf{A}_i$  y  $\mathbf{B}_i$  son matrices, cada una con dimensión  $n_i \times n_i$ , definidas en (5.29). Como resultado, la señal vectorial continua se modela por un proceso Gaussiano  $\mathbf{x}(t)$ , descrito por la ecuación diferencial vectorial y estocástica,

$$\frac{d\mathbf{x}}{dt} = \mathbf{A}(t)\mathbf{x}(t) + \mathbf{B}(t)\boldsymbol{\xi}(t), \quad \mathbf{x}(t_0) = \mathbf{x}_0$$

donde

$$\mathbf{A} = \begin{bmatrix} \mathbf{A}_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{A}_2 & \cdots & \mathbf{0} \\ \cdots & \cdots & \cdots & \cdots \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{A}_r \end{bmatrix}, \quad \mathbf{B} = \begin{bmatrix} \mathbf{B}_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{B}_2 & \cdots & \mathbf{0} \\ \cdots & \cdots & \cdots & \cdots \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{B}_r \end{bmatrix}, \quad \mathbf{x}(t) = \begin{bmatrix} x_1(t) \\ x_2(t) \\ \vdots \\ x_r(t) \end{bmatrix}, \quad \boldsymbol{\xi}(t) = \begin{bmatrix} \xi_1(t) \\ \xi_2(t) \\ \vdots \\ \xi_r(t) \end{bmatrix}$$

La excitación del sistema corresponde a proceso vectorial del tipo RBG  $\boldsymbol{\xi}(t)$  con valor medio  $E\{\boldsymbol{\xi}(t)\} = \mathbf{0}$  y función de correlación  $E\{\boldsymbol{\xi}(t)\boldsymbol{\xi}^\top(t-\tau)\} = \mathbf{R}_\xi(\tau)$ , que es una matriz de varianzas, simétrica y positiva definida, para los ruidos blancos componentes.

#### 5.3.4. Discretización de procesos continuos de Markov

El modelado de procesos continuos dinámicos de Markov en procesadores digitales se realiza mediante la discretización, sobre mallas de puntos determinados en el tiempo, a partir de las ecuaciones diferenciales que describen su comportamiento.

Sea un sistema dinámico, cuyo modelo se describe por la ecuación diferencial,

$$\frac{d\xi}{dt} = f(t, \xi) + g(t, \xi) d\nu(t), \quad \xi(t_0) = \xi_0, \quad t \in T$$

donde  $f$  y  $g$  son funciones determinísticas, continuas y diferenciables, que cumplen la condición de Lipschitz. Al considerar RBG,  $\eta(t)$ , a la entrada de un sistema, entonces su salida es un proceso escalar de Markov, determinado por la ecuación diferencial de Ito:

$$\frac{d\xi}{dt} = f(t, \xi) + g(t, \xi) d\nu(t), \quad \xi(t_0) = \xi_0, \quad t \in T \quad (5.36)$$

en el cual, las funciones  $g$  y  $f$  se consideran monótonas crecientes pero acotadas,

$$|f(t, \xi)|^2 + |g(t, \xi)|^2 \leq c^2(1 + |\xi|^2), \quad 0 < c < \infty, \quad t \in T$$

Las restricciones impuestas a los coeficientes (5.36) son suficientes para la existencia y unicidad de la solución  $\xi(t)$  en el intervalo  $t \in [0, T]$ , con lo cual a la ecuación estocástica diferencial (5.36) le corresponde la representación integral equivalente,

$$\xi(t) = \xi(t_0) + \int_{t_0}^t f(s, \xi(s)) ds + \int_{t_0}^t g(s, \xi(s)) ds d\nu(s) \quad (5.37)$$

La discretización del modelo del sistema se hace a partir de (5.36) o (5.37), para lo

cual, sobre un intervalo de análisis  $[0, T]$  se genera la malla de puntos uniformes de tiempo  $t_k = k\Delta$ , donde  $\Delta = T/n$ , es el paso de discretización. Luego, dada la condición inicial  $\xi_0 = \xi(0)$ , mediante el principio de estimación recurrente  $\tilde{\xi}_{v+1} = \tilde{\xi}_v + \theta_v(\tilde{\xi}_v, \Delta, \Delta v_v)$  para cada uno de los incrementos del proceso de Wiener,  $\Delta v_v = v(t_{v+1}) - v(t_v)$ , es necesario obtener los valores estimados  $\tilde{\xi}_v$ , que sean los más cercanos, para algún criterio dado a la solución verdadera  $\xi_v = \xi(t_v)$ . Esta cercanía de los valores de  $\xi_v$  y  $\tilde{\xi}_v$ , se da por el valor cuadrático medio del error  $\varepsilon$ , en la forma

$$\varepsilon = \max_{1 \leq v \leq n} \left\{ E\{(\tilde{\xi}_v - \xi_0)^2 | \xi_0\} \right\}^{1/2} \quad (5.38)$$

en el cual, las condiciones iniciales de las soluciones verdadera y aproximada coinciden, esto es,  $\xi_0 = \tilde{\xi}_{01}$ . Asumiendo la suavidad de las funciones  $f$  y  $g$  durante el intervalo de discretización, la expresión (5.37) corresponde a la ecuación iterativa de Euler,

$$\tilde{\xi}_{v+1} = \tilde{\xi} + f(v\Delta t, \tilde{\xi}_v)\Delta + g(v\Delta t, \tilde{\xi}_v)\Delta v \quad (5.39)$$

que en forma general, presenta un error del tipo  $\varepsilon = o(\sqrt{\Delta})$  [42].

La ecuación diferencial estocástica equivalente de Stratanovich para (5.39) tiene la siguiente forma:

$$d\xi = (f(t, \xi) - (1/2)g'_\xi(t, \xi)g(t, \xi))dt + g(t, \xi)dv, \quad \xi(0) = \xi_0 \quad (5.40)$$

donde  $g'_\xi(t, \xi) = \partial g(t, \xi)/\partial \xi$ . La equivalencia de (5.36) y (5.38) se entiende en el sentido, en que ambas expresiones, con iguales condiciones iniciales, conllevan a iguales resultados.

La representación integral equivalente de (5.40) tiene la forma.

$$\xi(t) = \xi(t_0) + \int_{t_0}^t (f(s, \xi(s)) + \frac{1}{2}g'_\xi(s, \xi(s))ds + \int_{t_0}^t g(s, \xi(s))dv(s) \quad (5.41)$$

La ecuación recursiva para (5.41) se expresa como

$$\tilde{\xi}_{v+1} = \tilde{\xi}_v + (f(v\Delta + \tilde{\xi}_v) - (1/2)g'_\xi(v\Delta, \tilde{\xi}_v))\Delta + g(v\Delta + \Delta/2, \tilde{\xi}_v + \Delta\tilde{\xi}_v/2)\Delta v + o(\Delta) \quad (5.42)$$

La aproximación de (5.42) se puede obtener si la función  $g(v\Delta + \Delta/2, \tilde{\xi}_v + \Delta\tilde{\xi}_v/2)$  se descompone en series de potencia sobre el punto  $(v\Delta, \tilde{\xi}_v)$  y se desprecian los términos, cuyo orden sea mayor a  $o(\Delta)$ . Luego (5.42) toma la forma

$$\begin{aligned} \tilde{\xi}_{v+1} = \tilde{\xi}_v + (f(v\Delta, \tilde{\xi}_v) - (1/2)g'_\xi(v\Delta, \tilde{\xi}_v)g(v\Delta, \tilde{\xi}_v))\Delta + g(v\Delta, \tilde{\xi}_v)\Delta v \\ + g'_\xi(v\Delta, \tilde{\xi}_v)(\Delta\tilde{\xi}_v/2)\Delta v + o(\Delta) \end{aligned}$$

Reemplazando en la anterior expresión el incremento,  $\Delta\tilde{\xi} = \tilde{\xi}_{v+1} - \tilde{\xi}_v$ , se obtiene que

$$\tilde{\xi}_{v+1} = \tilde{\xi}_v + f(v\Delta, \tilde{\xi}_v)\Delta + g(v\Delta, \tilde{\xi}_v)\Delta v + (1/2)g'_\xi(v\Delta, \tilde{\xi}_v)(\Delta v^2 + \Delta) + o(\Delta) \quad (5.43)$$

La expresión (5.43) se entiende como una aproximación estocástica mediante la serie de Taylor en cada intervalo de discretización, que en conjunto con la ecuación recursiva (5.41) permiten obtener la solución de la ecuación diferencial estocástica (5.36) de forma recurrente. Así por ejemplo, de (5.42) se obtiene una estructura más simple de cálculo,

$$\xi_{v+1} = \xi_v + f(t_v, \xi_v)\Delta + g(t_v, \xi_v)\Delta v \quad (5.44)$$

Sin embargo, la estructura recursiva obtenida de (5.43)

$$\xi_{v+1} = \xi_v + f(t_v, \xi_v)\Delta + g(t_v, \xi_v)\Delta v + (1/2)g'_\xi(t_v, \xi_v)g(t_v, \xi_v) + (\Delta v^2 - \Delta)$$

presenta menor de error de aproximación, si en (5.44) se asume  $g(t, \xi) = (N/2)^{1/2}$ :

$$\xi_{v+1} = \xi_v + f(t_v, \xi_v) + \sigma_v \eta_v \quad (5.45)$$

donde  $\eta_v$  es RBG con  $E\{\eta_i \eta_j\} = \delta_{ij}$ ,  $\sigma_v^2 = Ng^2(t_v, \xi_v)/2$ . La expresión (5.45) es un algoritmo recurrente de modelado de sucesiones de Markov  $\{\xi_v\}$ , con condición inicial  $\xi_0$ .

**Ejemplo 5.14.** Sea el proceso Gaussiano simple de Markov

$$\frac{d\xi}{dt} = -\alpha\xi + \gamma\eta(t) \quad (1)$$

La correspondiente ecuación recurrente, a partir de (5.45), toma la forma

$$\xi_{v+1} = \xi_v - \alpha\xi_v\Delta + \sigma_v\eta_v, \quad \sigma_v^2 = N\gamma^2\Delta/2 \quad (2)$$

Sin embargo otra forma de discretización de (1) se puede obtener de la solución general

$$\xi(t) = \xi(t_0)\exp(-\alpha(t-t_0)) + \gamma \int_{t_0}^t \exp(-\alpha(t-\tau))\eta(\tau)d\tau$$

con lo cual se obtiene

$$\xi(t) = \exp(-\alpha\Delta)\xi_v + \sigma(\Delta)\eta_v \quad (3)$$

donde

$$\sigma^2(\Delta) = \gamma^2 \frac{N}{2} \int_0^\Delta \exp(-2\tau)d\tau = \frac{\gamma^2 N}{4\alpha} (1 - \exp(-2\alpha\Delta)) \quad (4)$$

La expresión (4) implica que la varianza cumple la relación,  $\alpha\delta^2/d\Delta = -2\alpha\sigma^2 + \gamma^2 N/2$ , con condición inicial  $\sigma^2(0) = 0$ . Las expresiones (3) y (4), a diferencia de (2), establecen la relación precisa entre  $\xi_{v+1}$  y  $\xi_v$ , para cualquiera que sea el valor de  $\Delta$ . Si el proceso de discretización se escoge de tal manera que  $\alpha\Delta \ll 1$ , entonces se pueden emplear la aproximaciones  $\exp(-\alpha\Delta) \approx 1 - \alpha\Delta$  y  $\exp(-2\alpha\Delta) \approx 2\alpha\Delta$ . Luego, ambas expresiones (2) y (3) coinciden, y por lo tanto, la relación (2) será más precisa, entre menor sea el paso  $\Delta$ .

Los resultados anteriores se pueden generalizar para el caso de la ecuación diferencial lineal dada en forma vectorial con parámetros constantes,

$$\dot{\boldsymbol{\xi}} = \mathbf{A}\boldsymbol{\xi}(t) + \boldsymbol{\eta}(t) \quad (5.46)$$

donde  $\boldsymbol{\xi}(t)$  es el proceso vectorial,  $\mathbf{A}$  es una matriz cuadrada y  $\boldsymbol{\eta}(t)$  es el vector RBG con valor medio 0 y matriz de correlación,  $\mathbf{Q}$ , simétrica y positiva semidefinida:

$$E\left\{\boldsymbol{\eta}(t_1)\boldsymbol{\eta}^\top(t_2)\right\} = \mathbf{Q}\delta(t_2 - t_1) \quad (5.47)$$

La solución general de la ecuación (5.46) tiene la forma

$$\boldsymbol{\xi}(t) = \boldsymbol{\Phi}(t - t_0)\boldsymbol{\xi}(t_0) + \int_{t_0}^t \boldsymbol{\eta}(\boldsymbol{\Phi}(t - \tau)\boldsymbol{\eta}_0(\tau))d\tau \quad (5.48)$$

siendo  $\boldsymbol{\Phi}(t)$  la matriz de transición, que cumple la ecuación:  $\dot{\boldsymbol{\Phi}} = \mathbf{A}\boldsymbol{\Phi}(t)$ , con condición inicial  $\boldsymbol{\Phi}(0) = \mathbf{I}$ . De la ecuación (5.48) y teniendo en cuenta que,

$$\boldsymbol{\Phi}(t) = \exp(\mathbf{A}t) = \mathbf{I} + \mathbf{A}t + (1/2)\mathbf{A}^2t^2 + \dots \quad (5.49)$$

se obtiene la expresión recursiva  $\boldsymbol{\xi}_{v+1} = \exp(\mathbf{A}\Delta)\boldsymbol{\xi}_v + \boldsymbol{\eta}_v$ . La matriz de correlación del vector ruido blanco Gaussiano se determina, a partir de la expresiones (5.47) y (5.48),

$$\begin{aligned} \sigma^2(\Delta) = E\{\boldsymbol{\eta}_v\boldsymbol{\eta}_v^\top\} &= E\left\{\left(\int_{v\Delta}^{(v+1)\Delta} \exp(\mathbf{A}((v+1)\Delta - \tau_1))\boldsymbol{\eta}(\tau_1)d\tau_1\right) \times \right. \\ &\quad \left. \times \left(\int_{v\Delta}^{(v+1)\Delta} \exp(\mathbf{A}((v+1)\Delta - \tau_2))\boldsymbol{\eta}(\tau_2)d\tau_2\right)^\top\right\} \end{aligned}$$

que al introducir el operador de la esperanza se tiene que

$$\begin{aligned} \sigma^2(\Delta) &= \int_0^\Delta \int_0^\Delta \exp(\mathbf{A}^\top\tau_1) E\left\{\boldsymbol{\eta}((v+1)\Delta - \tau_1)\boldsymbol{\eta}^\top((v+1)\Delta - \tau_2)\right\} \times \\ &\quad \times \exp(\mathbf{A}^\top\tau_2) d\tau_1 d\tau_2 = \int_0^\Delta \exp(\mathbf{A}\tau) \mathbf{Q} \exp(\mathbf{A}^\top\tau) d\tau \end{aligned} \quad (5.50)$$

luego, al multiplicar por la matriz  $\mathbf{A}$  resulta en

$$\mathbf{A}\sigma^2(\Delta) = \mathbf{A} \int_0^\Delta \exp(\mathbf{A}\tau) \mathbf{Q} \exp(\mathbf{A}^\top\tau) d\tau = \int_0^\Delta (\mathbf{A} \exp(\mathbf{A}\tau) d\tau) \mathbf{Q} \exp(\mathbf{A}^\top\tau)$$



En este caso, se puede emplear la integración por partes en forma matricial

$$\begin{aligned}
 \mathbf{A}\boldsymbol{\sigma}^2(\Delta) &= \int_0^\Delta d(\exp(\mathbf{A}\tau)) \mathbf{Q} \exp(\mathbf{A}^\top \tau) d\tau \\
 &= \exp(\mathbf{A}\tau) \mathbf{Q} \exp(\mathbf{A}^\top \tau) \Big|_0^\Delta - \int_0^\Delta \exp(\mathbf{A}\tau) d(\mathbf{Q} \exp(\mathbf{A}^\top \tau)) \\
 &= \exp(\mathbf{A}\Delta) \mathbf{Q} \exp(\mathbf{A}^\top \Delta) - \mathbf{I} \mathbf{Q} \mathbf{I} - \left( \int_0^\Delta (\mathbf{A}\tau) \mathbf{Q} \exp(\mathbf{A}^\top \tau) d\tau \right) \mathbf{A}^\top \\
 &= \frac{d\boldsymbol{\sigma}^2(\Delta)}{d\Delta} - \mathbf{Q} - \boldsymbol{\sigma}^2(\Delta) \mathbf{A}^\top
 \end{aligned}$$

Con lo cual, la matriz  $\boldsymbol{\sigma}^2(\Delta)$  cumple la ecuación

$$\frac{d\boldsymbol{\sigma}^2(\Delta)}{d\Delta} = \mathbf{A}\boldsymbol{\sigma}^2 + \boldsymbol{\sigma}^2 \mathbf{A}^\top + \mathbf{Q} \quad (5.51)$$

con condición inicial  $\boldsymbol{\sigma}^2(0) = 0$ . La matriz  $\boldsymbol{\sigma}^2(\Delta)$  es preferible representarla en forma,

$$\boldsymbol{\sigma}^2 = \mathbf{G} \mathbf{G}^\top \quad (5.52)$$

La matriz  $\mathbf{G}$  es triangular inferior y genera una forma cómoda para el modelado,

$$\boldsymbol{\xi}_{v+1} = \exp(\mathbf{A}\Delta) \boldsymbol{\xi}_v + \mathbf{G} \boldsymbol{\eta}_v = \boldsymbol{\Phi}(\Delta) \boldsymbol{\xi}_v + \mathbf{G} \boldsymbol{\eta}_v \quad (5.53)$$

donde  $E\{\boldsymbol{\eta}_v \boldsymbol{\eta}_\mu^\top\} = \mathbf{I} \delta_{\mu v}$ .

La expresión (5.53) es la generalización vectorial de la forma escalar (3) y permite el modelado del proceso  $\boldsymbol{\xi}(t)$  en tiempo discreto. El algoritmo (5.53) para discretización de procesos continuos de Markov incluye los siguientes pasos:

1. Dada la matriz  $\mathbf{A}$ , mediante métodos de cálculo numérico se resuelve la ecuación  $\dot{\boldsymbol{\Phi}} = \mathbf{A}\boldsymbol{\Phi}(t)$  y se halla la matriz  $\boldsymbol{\Phi}(\Delta) = \exp(\mathbf{A}\Delta)$ .
2. Conocidas  $\mathbf{A}$  y  $\mathbf{Q}$ , por métodos numéricos se resuelve (5.51) y se halla  $\boldsymbol{\sigma}^2(\Delta)$ .
3. De la matriz  $\boldsymbol{\sigma}^2(\Delta)$  se halla la matriz cuadrada  $\mathbf{G}$ .
4. A partir de (5.53), dado  $\boldsymbol{\xi}_v$  se estima el siguiente valor discreto de la sucesión  $\boldsymbol{\xi}_{v+1}$ , para lo cual se emplea RBG, generado por algún método convencional, en calidad de señal de entrada.

En algunos casos, por ejemplo, cuando se tienen procesos vectoriales  $\boldsymbol{\xi}(t)$  en dimensión baja o cuando la matriz  $\mathbf{A}$  es diagonal, es preferible hallar de forma analítica las matrices  $\boldsymbol{\Phi}(t)$  y  $\boldsymbol{\sigma}^2(t)$ . En particular, empleando la descomposición matricial por Taylor de la matriz  $\boldsymbol{\Phi}(t)$ , y asumiendo la diagonalidad de  $\mathbf{A}$  se obtiene  $\exp(\mathbf{A}t) = \text{trace}[\exp(at) \cdots \exp(a_n t)]$ .

Sin embargo, la descomposición de la misma matriz  $\Phi(t)$  es más cómodo realizarla mediante la transformada de Laplace,  $\widehat{\Phi}(p) = \mathcal{L}\{\Phi(t)\}$ , en la cual cada elemento es la imagen por Laplace del respectivo elemento de la matriz  $\Phi(t)$ . Entonces, de  $\dot{\Phi} = \mathbf{A}\Phi(t)$  se obtiene

$$(p\mathbf{I} - \mathbf{A}) = \widehat{\Phi}(p) = \mathbf{I}, \quad \mathcal{L}\{\Phi(t)\} = p\widehat{\Phi}(p) - \Phi(0) = p\widehat{\Phi}(p) - \mathbf{I}$$

si se tiene que el determinante  $\det(p\mathbf{I} - \mathbf{A}) \neq 0$ , entonces,

$$\Phi(p) = (p\mathbf{I} - \mathbf{A})^{-1} \quad (5.54)$$

que se emplea en el cálculo de la exponente matricial  $\xi(t) = \mathcal{L}^{-1}\{(p\mathbf{I} - \mathbf{A})\}$ . La expresión vectorial para el caso lineal del algoritmo de aproximación escalar, ver ec. (2) del ejemplo 5.14, se obtiene truncando la serie de Taylor para  $t \rightarrow \infty$ , tal que,  $\Phi(t) \simeq \mathbf{I} + \mathbf{A}(t)$ , luego, de (5.50) se obtiene  $\sigma^2(\Delta) \sim \mathbf{A}\Delta$  y la representación vectorial de (5.51) toma la forma

$$\xi_{v+1} = (\mathbf{I} + \mathbf{A}\Delta)\xi_v + \mathbf{\Gamma}_1\eta_v, \quad E\{\eta_\nu\eta_\mu^\top\} = \mathbf{I}\delta_{\mu\nu} \quad (5.55)$$

La matriz diagonal inferior  $\mathbf{\Gamma}_1$  se obtiene de la expresión

$$\mathbf{\Gamma}_1\mathbf{\Gamma}_1^\top = \mathbf{Q}\Delta \quad (5.56)$$

**Ejemplo 5.15.** Basados en el algoritmo de representación matricial del algoritmo de la ec. (2) del ejemplo 5.14, hallar el modelo de una señal de voz, dado por el sistema

$$\frac{d\xi_1}{dt} + \alpha\xi_1 = \frac{d\xi_2}{dt}, \quad \frac{d\xi_2}{dt} + \beta\xi_2 = \eta(t) \quad (1)$$

La representación de (1) en la forma generalizada (5.46) es la siguiente

$$\frac{d\xi_1}{dt} = -\alpha\xi_1 + \beta\xi_2 + \eta(2), \quad \frac{d\xi_2}{dt} = -\beta\xi_2 + \eta(t) \quad (2)$$

Al comparar (5.46) con (2) se obtienen las matrices

$$\mathbf{A} = \begin{bmatrix} -\alpha & -\beta \\ 0 & -\beta \end{bmatrix}, \quad \mathbf{Q} = \frac{N}{2} \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \quad (3)$$

Mediante (5.54), se halla la transformada de Laplace de la matriz  $\Phi(t)$ ,

$$\widehat{\Phi}(p) = (p\mathbf{I} - \mathbf{A})^{-1} = \begin{bmatrix} p + \alpha & \beta \\ 0 & \beta \end{bmatrix}^{-1} = \frac{1}{(p + \alpha)(p + \beta)} \begin{bmatrix} p + \beta & -\beta \\ 0 & p + \beta + \alpha \end{bmatrix} \quad (4)$$

Luego, la matriz de transición obtenida es

$$\Phi(t) = \mathcal{L}^{-1}\{\widehat{\Phi}(p)\} = \begin{bmatrix} \exp(-\alpha t) & \frac{\beta}{\alpha - \beta}(\exp(-\alpha t) - \exp(-\beta t)) \\ 0 & \exp(-\beta t) \end{bmatrix}$$

Basados en los valores (3) y (4), además teniendo en cuenta (5.50), se determinan los valores

de la matriz de varianza  $\sigma^2(\Delta)$  (siendo  $T(x) = 1 - \exp(-\Delta x)$ ):

$$\begin{aligned} d_{11} &= \frac{N_0}{2(\alpha - \beta)^2} \left( \frac{\alpha}{2} T(2\alpha) - \frac{\alpha\beta}{\alpha + \beta} T(\alpha + \beta) + \frac{\beta}{2} T(2\beta) \right) \\ d_{12} &= \frac{N_0}{2(\alpha - \beta)^2} \left( \frac{\alpha}{2} T(\alpha + \beta) - \frac{1}{1(\alpha + \beta)} T(2\beta) \right) \\ d_{22} &= \frac{1}{2\beta} T(2\beta) \end{aligned} \quad (5)$$

Por lo anterior, un proceso vectorial aleatorio de Markov  $\xi^T = \{\xi_1, \xi_2\}$  se modela mediante el algoritmo (5.53), si se descompone la matriz  $\sigma^2(\Delta)$  con elementos (5) de acuerdo con la expresión (5.52). El algoritmo de aproximación (5.55) se puede obtener de dos formas: empleando la expresión (5.56) o hallando la primera aproximación, cuando  $\Delta \rightarrow 0$ , para la expresión (5). En ambos casos se obtiene el mismo resultado.

Luego

$$\Phi(\Delta) \sim \begin{bmatrix} 1 - \alpha\Delta & 1 - \beta\Delta \\ 0 & 1 - \beta\Delta \end{bmatrix}, \quad \sigma^2(\Delta) \sim \frac{N\Delta}{2} \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$$

El reemplazo de los valores respectivos implica que la solución de la ecuación  $\Gamma_1 \Gamma_1^T = \sigma^2(\Delta)$  corresponde a la matriz

$$\Gamma_1 = \left( \frac{N\Delta}{2} \right)^{1/2} \begin{bmatrix} 1 & 0 \\ 1 & 0 \end{bmatrix}$$

El algoritmo de aproximación (5.53) para el ejemplo en análisis tiene la forma

$$\xi_{v+1} = \begin{bmatrix} 1 - \alpha t & \beta \Delta t \\ 0 & 1 - \beta \Delta t \end{bmatrix} \xi_v + \left( \frac{N\Delta}{2} \right)^{1/2} \begin{bmatrix} 1 & 0 \\ 1 & 0 \end{bmatrix} \eta_v$$

**Ejemplo 5.16.** Sea un proceso bicomponente Gaussiano de Markov en la forma

$$\frac{\xi_1}{dt} = -\alpha_1 \xi_1 - \beta_1 \xi_2 + \eta_1(t), \quad \frac{\xi_2}{dt} = -\alpha_2 \xi_1 - \beta_2 \xi_2 + \eta_2(t)$$

$$\text{donde } E\{\eta_i(t)\} = 0, \quad E\{\eta(t_1)\eta^T(t_2)\} = \begin{bmatrix} \frac{N}{2} & r \frac{\sqrt{N_1 N_2}}{2} \\ r \frac{\sqrt{N_1 N_2}}{2} & \frac{N_2}{2} \end{bmatrix}.$$

La matriz  $\mathbf{A}$ , en este caso, de determina como

$$\mathbf{A} = \begin{bmatrix} -\alpha_1 & -\beta_1 \\ -\alpha_2 & -\beta_2 \end{bmatrix}$$

El cálculo de la matriz  $\Phi$  se puede realizar mediante el método de diagonalización. En estos casos, inicialmente se calculan los valores propios  $\lambda_{1,2}$  de la matriz  $\mathbf{A}$ , como la solución de la ecuación  $\det(\lambda \mathbf{I} - \mathbf{A}) = 0$ :

$$\lambda_{1,2} = -(\alpha_1 + \beta_2 \mp \omega)/2, \quad \omega = ((\alpha_1 - \beta_2)^2 + 4\alpha_2\beta_1)^{1/2}$$

asumiendo que  $\alpha_{1,2} \in \mathbb{R}$ , entonces existe una matriz regular  $\mathbf{C}$  que transforme a  $\mathbf{A}$  a la forma

diagonal, esto es,

$$\mathbf{C}\mathbf{A}\mathbf{C}^{-1} = \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix} = \mathbf{R} \quad (6)$$

Los elementos de la matriz  $\mathbf{C}$  se hallan de la ecuación obtenida de la multiplicación de la igualdad (6) por  $\mathbf{C}$ :

$$\mathbf{A} = \mathbf{R}\mathbf{C}, \Rightarrow \mathbf{C} = \begin{bmatrix} \nu/2\beta_1 & 1 \\ 1 & -\nu/2\alpha_2 \end{bmatrix}, \mathbf{C}^{-1} = 2 \begin{bmatrix} \nu\beta_1/\delta & 2\alpha_2\beta_1/\delta \\ 2\alpha_2\beta_1/\delta & -\nu\alpha_2/\delta \end{bmatrix}$$

donde  $\nu = \alpha_1 - \beta_2 - \omega$ ,  $\delta = \nu^2 + \alpha_2\beta_1$ .

De (6), se tiene que  $\mathbf{A} = \mathbf{C}^{-1}\mathbf{R}\mathbf{C}$ , por lo tanto,

$$\mathbf{A}^n = (\mathbf{C}^{-1}\mathbf{R})(\mathbf{C}^{-1}\mathbf{R}\mathbf{A}) \dots (\mathbf{A}^{-1}\mathbf{R}\mathbf{C}) = \mathbf{C}^{-1}\mathbf{R}^n\mathbf{C}$$

El empleo de la descomposición (5.49) de la matriz

$$\Phi(t) = \exp(\mathbf{A}t) = \mathbf{I} + \mathbf{A}t + (1/2)\mathbf{A}^2t^2 + \dots + \mathbf{C}^{-1}\mathbf{I}\mathbf{C} + \mathbf{C}^{-1}\mathbf{R}t\mathbf{C} + \dots \quad (7)$$

$$= \mathbf{C}^{-1}(\mathbf{I} + \mathbf{R}t + (1/2)\mathbf{R}^2t^2 + \dots)\mathbf{C} = \mathbf{C}^{-1}\exp(\mathbf{R}t)\mathbf{C} \quad (5.57)$$

Por cuanto la matriz  $\mathbf{R}$  es ortogonal, entonces  $\exp(\mathbf{A}t) = \begin{bmatrix} \exp(\lambda_1 t) & 0 \\ 0 & \exp(\lambda_2 t) \end{bmatrix}$

Teniendo en cuenta las expresiones de  $\mathbf{C}$  y  $\mathbf{C}^{-1}$ , además la relación (7), se halla que

$$\phi(\Delta) = \begin{bmatrix} \nu^2 \exp(\lambda_1 \Delta) + 4\lambda_2 \beta_1 \exp(\lambda_2 \Delta) & 2\nu \lambda_2 (\exp(\lambda_1 \Delta) - \exp(\lambda_2 \Delta)) \\ 2\nu \beta_1 (\exp(\lambda_1 \Delta) - \exp(\lambda_2 \Delta)) & 4\lambda_2 \beta_1 \exp(\lambda_1 \Delta) + \nu^2 \exp(\lambda_2 \Delta) \end{bmatrix}$$

Los elementos de la matriz de varianza  $\sigma^2(\Delta)$  se muestran en [41].

## Problemas

**Problema 5.17.** La ecuación diferencial de un generador de señales senoidales tiene la forma

$$\dot{a} - ba \left(1 - \frac{a^2}{4}\right) = c\xi(t)$$

donde  $a$  es la amplitud,  $b, c = \text{const}$  y  $\xi(t)$  son las fluctuaciones que generan cambios en la amplitud. Hallar la ecuación diferencial lineal con relación al incremento relativo de amplitud y su solución estacionaria, empleando el método de linealización determinística. Asumir  $E\{\xi(t)\} = 0$  y  $\sigma_a \ll a_0$ .

**Problema 5.18.** Hallar el respectivo modelo de la señal de un proceso  $\xi(t)$ , a la salida de un circuito de resonancia, dado por la siguiente ecuación lineal diferencial de segundo orden:

$$d^2\xi/dt^2 + 2\alpha d\xi/dt + \omega_0^2\xi = \omega_0^2\eta(t), \quad \omega_0 \gg \alpha$$

# Capítulo 6

## Detección y filtración de señales aleatorias

Dada una mezcla de la señal útil  $\xi$  y alguna perturbación  $\eta$  presente en ella, el objetivo de la filtración consiste en obtener la estimación, bien sea de la misma señal útil en general,  $\tilde{\xi}$ , o bien de alguno de sus parámetros o características de interés,  $\xi(\tilde{\theta})$ , con estructura aleatoria. De otra manera, la filtración implica buscar la transformación de la trayectoria  $\mathbf{x}(t) \in \xi$  en análisis a fin de obtener la mejor estimación, por un criterio dado, de los valores de un proceso aleatorio en un momento determinado del tiempo del intervalo de observación. Además de la estimación del mismo proceso  $\xi$ , en la práctica, a partir de las respectivas observaciones  $\mathbf{x} \in \xi$ , dadas sobre un intervalo de tiempo de análisis, tiene sentido la estimación de alguna transformación lineal del proceso aleatorio, por ejemplo, el desplazamiento en el tiempo, la diferenciación o integración múltiple o la combinación de estas transformaciones.

Aunque en la mayoría de los casos, los sistemas óptimos de proceso son no lineales. No obstante, el desarrollo de los métodos de filtración lineal tiene mayor importancia, por cuanto su aplicación presenta mejor sentido práctico. De otra parte, debido a que la condición de linealidad no implica la condición de realización física (causalidad del sistema), es usual para la síntesis de filtros el empleo de la solución secuencial: primero, la comprobación de la condición de linealidad y luego, la de causalidad.

Como criterio de calidad en la estimación del valor de la señal aleatoria, en algunos casos, se escoge el valor medio de la potencia del error para un ensamble dado de trayectorias, esto es, el valor cuadrático medio de la desviación estándar de la estimación del valor estimado. En otros casos, el criterio de calidad corresponde a la relación de las potencias de la señal contra la de perturbación.

El análisis de filtros no lineales se puede considerar de diversas maneras; cada una de ellas acopladas a la naturaleza de no linealidad del proceso. En general, entre las aproximaciones más conocidas, descritas en el numeral §5.2.1, están las relacionadas con la caracterización de los sistemas no lineales mediante la suma infinita de los integrales de Volterra. Otra aproximación consiste en la representación, dadas las restricciones pertinentes, en forma de procesos de Markov.

## 6.1. Métodos de detección

Sea el conjunto de señales de información,  $\{\mathbf{x}_k(t) : k = 1, \dots, n\}$  con estructura conocida, bien sea de naturaleza determinística o aleatoria, y de las cuales solo puede existir una señal,  $\mathbf{x}_k(t)$ , durante un intervalo dado de observación  $T$ , que además se asume perturbada por alguna trayectoria de la distorsión aleatoria  $\boldsymbol{\eta}(t)$ .

A efectos de análisis, la perturbación  $\boldsymbol{\eta}(t)$  se considera RBG aditivo, con lo cual, la trayectoria sobre la cual se realiza la detección tiene la forma:

$$\mathbf{y}(t) = \mathbf{x}_k(t) + \boldsymbol{\eta}(t), \quad t \in T \quad (6.1)$$

De tal manera, que pueden generarse las respectivas hipótesis sobre cuál de las  $k$  posibles señales está presente en el intervalo dado de observación.

La tarea de detección, de acuerdo con el número  $k$  de posibles señales de información, puede ser binaria, ( $k = 2$ ), o bien múltiple, ( $k > 2$ ). Sobre el mismo carácter de las señales de información, se pueden diferenciar las siguientes tres tareas:

- (a). Detección de las señales determinísticas con parámetros conocidos (sistemas sincrónicos de comunicación digital).
- (b). Detección de señales determinísticas con parámetros desconocidos (detección de radar, sistemas asincrónicos de comunicación digital).
- (c). Detección de señales de información con estructura aleatoria (detección sísmica, radioastronomía, radioprospección).

Sea un conjunto de señales  $\mathbf{x}_k(t)$ ,  $k = 1, \dots, N$ , dado el modelo de observación (6.1), para el cual se introduce la FDP condicional con dimensión múltiple,  $p(\mathbf{y}|\mathbf{x}_k)$  de la variable  $\mathbf{y}$ , bajo la condición de que esté presente la señal de información  $\mathbf{x}_k$ , en el intervalo dado de observación. La toma de decisión sobre cuál de las posibles señales está presente, implica que todo el espacio  $\mathfrak{G}$  conformado por las señales de información sea dividido de forma determinada en  $m$  subespacios,  $\{G_i \subset \mathfrak{G} : i = 1, \dots, N\}$ . De tal manera, que al tenerse  $\mathbf{y}(t) \in G_i$ ,  $t \in T$ , entonces, se toma la decisión de que está presente la señal  $\mathbf{x}_i(t)$ .

Sin embargo, debido a la interacción del ruido  $\boldsymbol{\eta}(t)$ , el verdadero valor de la medida de la señal presente  $\mathbf{x}_k(t)$  se altera, como se muestra en la Figura 6.1, luego, pueden ocurrir dos casos extremos en la toma de decisión:

- Primero, que la perturbación  $\boldsymbol{\eta}_1(t)$  cambie el valor real de medida, pero la señal  $\mathbf{x}_k(t)$  siga reflejándose, mediante  $\mathbf{y}_1(t)$ , en su respectivo subespacio  $G_k$ , con lo que la decisión se toma correctamente sobre cuál de las señales fue detectada.
- Segundo, que la trayectoria  $\boldsymbol{\eta}_1(t)$  altere la medida, de tal manera que la señal  $\mathbf{x}_k(t)$ , reflejada en  $\mathbf{y}_2(t)$ , traspase la frontera  $\gamma_{ki}$  y se ubique en un subespacio diferente, esto es,  $\mathbf{x}_k \rightarrow G_i$ , para todo  $i = 1, \dots, N$ ,  $i \neq k$ ; por lo tanto, se genera un error en la toma de decisión durante la detección de la respectiva señal.

Es evidente, que la conformación de los subespacios y, en particular, el cambio de las fronteras influyen sobre la probabilidad de error de la detección de las señales de información. Así por ejemplo, si en vez de tomar la frontera  $\gamma_{km}$ , dada la observación  $y_1$ , entonces la detección de  $x_k(t)$  es incorrecta, por cuanto se toma la decisión  $G_m$ . Por el contrario, para la trayectoria  $y_2$ , al cambiar la frontera hasta  $\gamma_{ki}$  el espacio de decisión  $G_i$  se amplía y la detección de  $x_k$  ya se convierte en correcta.

La FDP condicional de detección correcta se da por la integral múltiple:

$$p(x_i|x_i) = \int_{G_i \in \mathcal{G}} p(y|x_i) dy \quad (6.2)$$

Mientras, la FDP condicional complementaria de error tiene la forma

$$p(x_j|x_i) = \int_{G_i} p(y|x_i) dy, \quad \forall j = 1, \dots, N, \quad j \neq i \quad (6.3)$$

De esta manera, dado un criterio de decisión, la mejor división de los subespacios de las señales de información (toma de decisión óptima de detección) se realiza mediante los métodos de teoría de toma de decisión estadística, descritos en el numeral §2.3.

### 6.1.1. Detección bayesiana de señales

Sea la función de pérdida  $f_c\{\varepsilon\}$  en forma de los valores  $c_{ji}$ , que resultan de la toma incorrecta de la decisión de detección de la señal  $x_i$ , cuando realmente se tenía  $x_j$ . Las funciones de pérdida asumen  $c_{ij} = 0$ , sin embargo, se puede generalizar su definición y considerar el caso cuando  $c_{ij} \neq 0$ , asumiendo el sentido de *función de costos* de decisión.

La función de riesgo condicional, de (2.81), se determina para la detección de  $x_i$  como

$$r_g(x_i) = \sum_{n=1}^N c_{ij} p(x_j|x_i)$$

Dada la FDP a priori de aparición,  $p(x_i)$ , para la señal  $x_i$ , entonces el riesgo medio (2.86), cuando se asume la presencia de una sola señal de información en el intervalo de análisis, se determina como

$$R_g(x_i) = \sum_{n=1}^N r_g(x_i) p(x_i) = \sum_{n=1}^N \sum_{n=1}^N c_{ij} p(x_i) p(x_j|x_i)$$

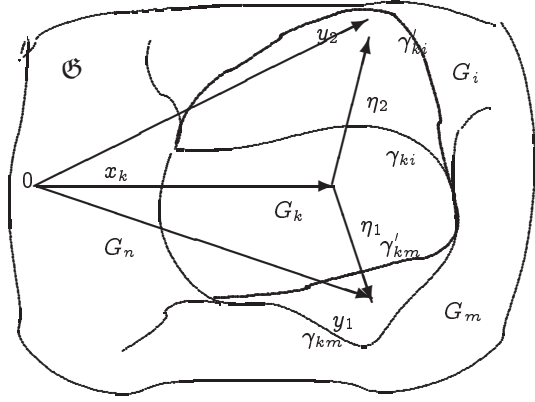


Figura 6.1. Espacios de decisión

Teniendo en cuenta (6.2) y (6.3) se obtiene

$$R_g(x_i) = \sum_{n=1}^N \sum_{n=1}^N c_{ij} p(x_i) \int D_j p(y|x_i) dy \quad (6.4)$$

El método de detección se considera óptimo, de acuerdo al criterio de riesgo medio, cuando el valor de (6.4) se hace mínimo, cuyas variables de optimización son las características de las señales, en particular la estructura y parámetros de los operadores de transformación de las señales mismas, que conforman las fronteras de los respectivos subespacios de toma de decisión.

### Detección binaria

Sea una observación,  $y(t)$ , que corresponde al modelo (6.1), la cual se representa por una sucesión de valores  $y[n]$ ,  $n = 0, \dots, N-1$ , que corresponde a un vector,  $\mathbf{y}$ , con dimensión  $N$ . Sean las hipótesis  $H_0$  y  $H_1$ , conformadas para los casos de aparición de las señales  $x_0$  y  $x_1$ , respectivamente, con lo cual se pueden tener los siguientes resultados de decisión:

1. Se acepta  $H_0$  y  $H_0$  es cierta, entonces se asigna el costo de decisión  $c_{00}$
2. Se acepta  $H_1$ , pero  $H_1$  es cierta,  $c_{10}$
3. Se acepta  $H_1$  y  $H_1$  es cierta,  $c_{11}$
4. Se acepta  $H_0$ , pero  $H_1$  es cierta,  $c_{01}$

Las decisiones 1 y 3 son correctas, mientras las decisiones 2 y 4 resultan incorrectas y generan el correspondiente error.

Definidas las probabilidades a priori  $P_0$  y  $P_1$ , que corresponden a la probabilidad de aparición de las señales  $x_0$  y  $x_1$ , respectivamente, el riesgo medio (6.4) resulta en

$$R_g = c_{00}p(x_0)p(x_0|x_0) + c_{10}p(x_0)p(x_1|x_0) + c_{11}p(x_1)p(x_1|x_1) + c_{01}p(x_1)p(x_0|x_1) \quad (6.5)$$

El valor de la función de costos se define de tal manera que, los coeficientes de error sean mayores que cero,  $c_{10} > 0, c_{01} > 0$ , mientras para los coeficientes de decisión acertada se tenga,  $c_{00} \leq 0, c_{11} \leq 0$ , con lo cual se cumple que,

$$c_{10} > c_{00}, c_{01} > c_{11}$$

La función de decisión consiste en la conformación, a partir del espacio total de decisión y dada la función de transformación del espacio de observaciones,  $D = g(y)$ , para los subespacios de decisión  $D_0$  y  $D_1$ , que corresponden a las respectivas hipótesis:  $H_0$  y  $H_1$ .

Como antes se dijo, la regla de decisión óptima consiste en la determinación de la frontera  $\gamma$ , que minimice el riesgo medio (6.4). Por cuanto, para el subespacio  $D_i$ ,  $i = 0, 1$ , se toma



la hipótesis  $H_i$ , entonces

$$p(x_0|x_0) = \int_{D_0} p(y|x_0) dy \quad (6.6a)$$

$$p(x_1|x_0) = \int_{D_1} p(y|x_0) dy \quad (6.6b)$$

$$p(x_1|x_1) = \int_{D_1} p(y|x_1) dy \quad (6.6c)$$

$$p(x_0|x_1) = \int_{D_0} p(y|x_1) dy \quad (6.6d)$$

donde  $p(y|x_0)$  y  $p(y|x_1)$  son la FDP condicionales de aparición de la trayectoria  $y(t)$ , cuando se asume la presencia de las señales de información,  $x_0$  y  $x_1$ , respectivamente.

Las integrales de ambas FDP, tomadas por los respectivos subespacios de decisión,  $D_0$  y  $D_1$ , determinan la validez de las correspondientes hipótesis,  $H_0$  y  $H_1$ . Reemplazando en (6.5), las expresiones (6.5) y (6.6d), se obtiene el valor de riesgo

$$\begin{aligned} R = c_{00}p(x_0) \int_{D_0} p(y|x_0) dy + c_{10}p(x_0) \int_{D_1} p(y|x_0) dy \\ + c_{11}p(x_1) \int_{D_1} p(y|x_1) dy + c_{01}p(x_1) \int_{D_0} p(y|x_1) dy \end{aligned}$$

Las relaciones (6.2) y (6.3) son complementarias, en el sentido en que,

$$\int_{D_1} p(y|x_i) dy = 1 - \int_{D_0} p(y|x_i) dy, \quad i = 0, 1 \quad (6.7)$$

luego, el riesgo se expresa en función de un solo subespacio de decisión, por ejemplo,  $D_0$ ,

$$\begin{aligned} R = p(x_0)c_{10} + p(x_0)c_{11} \\ + \int_{D_0} (p(x_1)(c_{01} - c_{11})p(y|x_1) - (p(x_0)(c_{10} - c_{00}))p(y|x_0)) dy \quad (6.8) \end{aligned}$$

Los dos primeros términos en (6.8) corresponden a los valores fijos de pérdida, que no dependen de la selección de las fronteras del subespacio de decisión  $D_0$ . Por lo tanto, la minimización del riesgo  $R$  significa configurar el subespacio  $D_0$ , de tal manera que se incluyan únicamente los valores negativos (valores de pérdida) de la integral en la expresión (6.8), esto es, de la condición:

$$p(x_1)(c_{01} - c_{11})p(y|x_1) - p(x_0)(c_{10} - c_{00})p(y|x_0) < 0$$

que se puede escribir en la forma

$$\frac{p(y|x_1)}{p(y|x_2)} < \frac{p(x_0)(c_{10} - c_{00})}{p(x_1)(c_{01} - c_{11})}$$

De lo anterior, la función de decisión toma la forma

$$\frac{p(y|x_1)}{p(y|x_2)} \underset{H_0}{\overset{H_1}{>}} \frac{p(x_0)(c_{10} - c_{00})}{p(x_1)(c_{01} - c_{11})} = \gamma \quad (6.9)$$

La parte derecha de la expresión (6.9) es la función de verosimilitud,

$$\Lambda(y) \triangleq \frac{p(x_0)(c_{10} - c_{00})}{p(x_1)(c_{01} - c_{11})}$$

que implica en (6.9) que es más verosímil la hipótesis con la mayor FDP a posteriori, para una trayectoria dada de  $y(t)$ .

La parte derecha de (6.9) corresponde a un valor fijo  $\gamma$ , calculado a partir de la relación entre las probabilidades de aparición de las señales  $x_0$  y  $x_1$  y los respectivos costos de decisión, que se determina como

$$\gamma = \frac{p(x_0)(c_{10} - c_{00})}{p(x_1)(c_{01} - c_{11})}$$

En general, el riesgo medio es uno de los criterios de decisión más generalizados, pero su empleo implica una gran cantidad de información sobre los procesos de análisis, que en la práctica no siempre está disponible.

### Detección del observador ideal

Sean los coeficientes de la función de costos iguales y definidos de la forma:

$$c_{ij} = \begin{cases} 0, & i = j \\ 1, & i \neq j \end{cases}$$

con lo cual se asume que todos los errores de decisión conllevan a un mismo costo y, por lo tanto, el riesgo medio corresponde a la probabilidad total de aparición del error

$$R = P_e = \sum_{i=1}^N \sum_{\substack{j=1 \\ j \neq i}}^N p(x_i) \int_{D_j} p(y|x_i) dy \quad (6.10)$$

El criterio (6.10) determina el mínimo valor medio de errores, mientras la probabilidad de detección correcta sea la máxima, que se alcanza cuando la decisión de que la señal detectada pertenece al subespacio  $D_i$ , cumple la condición,

$$p(y|x_i)p(x_i) > p(y|x_j)p(x_j), \quad \forall j \neq i \quad (6.11)$$

En esencia, dadas las  $N - 1$  condiciones (6.11), se tiene la siguiente regla de detección:

$$\max_j \{p(x_j)p(y|x_i)\} = p(x_i)p(y|x_i) \quad (6.12)$$

esto es, la decisión sobre cuál fue la señal detectada recae en aquella que tenga la máxima FDP a priori.

### Detección del máximo de probabilidad a posteriori

Debido a que  $p(x_i)p(y|x_i) = p(x_i|y)p(y)$ , donde  $p(x_i|y)$  es la FDP a posteriori de que se tiene la señal  $x_i$ , cuando se mide la observación  $y$ , mientras,  $p(y)$  es la densidad marginal de la señal medida, entonces, en concordancia con el teorema de Bayes (2.4), se tiene,

$$p(x_i|y) = \frac{p(x_i)p(y|x_i)}{\sum_{i=1}^N p(x_i)p(y|x_i)} \quad (6.13)$$

luego, el algoritmo de detección óptimo, en este caso, tiene la forma

$$\max_j \{p(x_i|y)\} = p(x_i|y) \quad (6.14)$$

La comparación de los algoritmos (6.12) y (6.14) muestra que ambos conllevan a la misma decisión óptima, y por lo tanto, los correspondientes criterios son equivalentes. Además, ambos exigen la misma información a priori del proceso en análisis.

### Detección minimax

En caso de desconocer las probabilidades a priori de aparición de las señales, el riesgo (6.4) se puede calcular sobre la peor densidad de probabilidad,  $p(x_i)$ ,  $i = 1, \dots, N$ , que se asume de tal manera que se obtenga el mayor valor del riesgo medio, luego,

$$R = \min_{D_j} \max_{p(x_i)} \sum_{i=1}^N \sum_{j=1}^N c_{ij} p(x_i) \int_{D_j} p(z|x_i) dz$$

### Detección de Neymann-Pearson

Existen casos en los cuales se tiene una clara asimetría en las pérdidas de decisión. En el caso particular de la detección binaria, el criterio minimiza uno de los dos tipos de error indicados en el ejemplo 2.27.

En la práctica, tiene sentido brindar la mínima probabilidad de error del tipo 1,

$$p(x_0|x_1) = \int_{D_0} p(y|x_1) dy$$

para un valor dado probabilidad de error tipo 2,

$$p(x_1|x_0) = \int_{D_1} p(y|x_0) dy$$

con lo cual, el algoritmo de detección (6.14), para el caso de detección binaria, se simplifica significativamente hasta la expresión  $p(x_1|y) > p(x_0|y)$ , que al tomar en cuenta (6.13) puede describirse de cualquiera de las siguientes dos formas:

$$\begin{aligned} p(x_1)p(y|x_1) &> p(x_0)p(y|x_0) \\ p(y|x_1)/p(y|x_0) &> p(x_0)/p(x_1) \end{aligned}$$

Esto es,  $\Lambda(y) = \gamma$ , donde  $\gamma = p(x_0)/p(x_1)$ . El umbral  $\gamma$  no es conocido, pero se escoge a partir de un valor asumido para el error  $\alpha_2$  del tipo 2. Ambas FDP de error se expresan en función de la relación de verosimilitud, haciendo el cambio de variables,  $y \rightarrow \Lambda$ ,

$$p(y|x_1)dy = p(\Lambda|x_1)d\Lambda, \quad p(y|x_0)dy = p(\Lambda|x_0)d\Lambda$$

En este caso, el espacio de decisión  $D$  se transforma en un eje de valores,  $\Lambda$ , en el cual el valor  $\gamma$  corresponde a la frontera entre los espacios de decisión, por lo tanto,

$$\begin{aligned} p(x_0|x_1) &= \int_{D_0} p(y|x_1)dy = \int_0^\gamma p(\Lambda|x_1)d\Lambda \\ p(x_1|x_0) &= \int_{D_1} p(y|x_0)dy = \int_\gamma^\infty p(\Lambda|x_0)d\Lambda \end{aligned}$$

El valor  $\gamma$ , entonces, se puede determinar dada la condición  $\int_\gamma^\infty p(\Lambda|x_0)d\Lambda = \alpha_2$ . Como resultado, el detector óptimo por Neymann-Pearson realiza el algoritmo

$$\Lambda > \gamma \tag{6.15}$$

Si la condición (6.15) es cierta, entonces, el algoritmo asume la detección de  $x_1(t)$ .

### Detección por medidas de información

En este caso, el criterio analiza la cantidad de información en la detección, para lo cual se emplea la medida (2.57) que relaciona la información mutua de una señal dada con respecto a los demás, sobre su entropía, (2.58). En el caso binario, este criterio de detección escoge la señal con mayor relación,

$$R = I(x_j, x_i)/H(x_i) = 1 - H(x_i|x_j)/H(x_i), \quad i \neq j, i, j = 0, 1$$

Cabe anotar, que del ejemplo 2.18, dados dos sucesos con probabilidades respectivas  $p_1$

y  $p_2$ , la mayor entropía se obtiene para  $p_1 = p_2$ :

$$H(x_1|x_0) = -(p_1 \ln p_1 + (1 - p_1) \ln (1 - p_1))$$

En la práctica, el criterio de detección informativo conlleva a los mismos resultados que los criterios (6.12) y (6.13).

### 6.1.2. Detección de máxima verosimilitud

Si en el algoritmo (6.12), se asume la FDP uniforme para todas las señales,  $p(x_j) = 1/N$ , entonces el criterio se expresa solamente en términos de la función de verosimilitud,

$$p(y|x_i) = \max_j \{p(y|x_j)\} \quad (6.16)$$

El criterio de detección (6.16) ha obtenido mayor aceptación debido a la simplicidad relativa de implementación, que exige espacios de observación no muy grandes.

De otra parte, teniendo en cuenta (6.9) el criterio de Bayes se puede expresar en términos del criterio de máxima verosimilitud

$$\Lambda(y) \underset{H_0}{\overset{H_1}{>}} \gamma \quad (6.17)$$

Los valores de las densidades a priori y de los costos influyen solamente en el cálculo del umbral, mas no en la estructura del algoritmo, que se basa en la estimación de  $\Lambda(y)$ . En este sentido la sintonización de los valores de los costos y probabilidades de generación de las señales, que frecuentemente se hace de forma heurística, no influye en la forma de toma de decisión, y más bien ocurre la sintonización fina de la ubicación del umbral.

Debido a que ambas partes de la igualdad (6.17) son positivas, entonces en vez de tomar directamente la relación de verosimilitud, se emplea su logaritmo, que es una función monótona. Así, el criterio (6.17) toma la forma

$$\ln \Lambda(y) \underset{H_0}{\overset{H_1}{>}} \ln \gamma \quad (6.18)$$

que es adecuada, debido a que en la práctica la mayoría de FDP  $p(y|x_i)$  son exponenciales.

**Ejemplo 6.1.** Desarrollar el criterio de máxima verosimilitud para la detección binaria de las señales  $x_0(t)$  y  $x_1(t)$ , afectadas por el ruido blanco Gaussiano  $\eta(t)$  con parámetros  $\mathcal{N}(0, \sigma_\eta)$ ; asumiendo que el ancho de banda del espectro de las señales es  $\Delta f$ .

Sea la trayectoria del ruido  $\eta(t)$  que se discretiza, con periodo de muestreo  $\Delta t = 1/2\Delta f$ , y conforma la sucesión  $\{\eta_i : i = 1, m\}$ , en el cual se considera que todos sus valores son variables aleatorias independientes. Por lo tanto, la correspondiente FDP de dimensión múltiple Gaussiana se puede representar en la forma,

$$P_\eta(\eta_1, \eta_2, \dots, \eta_m) = \frac{1}{(2\pi\sigma_\eta^2)^{m/2}} \prod_{k=1}^m \exp(-\eta_k^2/2\sigma_\eta^2) = \frac{1}{(2\pi\sigma_\eta^2)^{m/2}} \exp\left(-\frac{1}{2\sigma_\eta^2} \sum_{k=1}^m \eta_k^2\right)$$

donde  $m = T/\Delta t = 2T\Delta f$ , siendo  $T$  el intervalo de análisis para la detección.

La FDP condicional  $p(y|x_i)$  de aparición de la trayectoria de la medida,  $y(t) = x_i(t) + \eta(t)$ , corresponde a la densidad de probabilidad,  $P_\eta(y(t) - x_i(t))$ , de la observación del ruido  $\eta(t) = y(t) - x_i(t)$ , que conforman la sucesión de valores discretizados,

$$\eta_k \neq \eta(t_k) = y(t_k) - x_i(t_k) = \Delta y_i[k]$$

que le corresponde la FDP con dimensión múltiple de la forma

$$p(y|x_i) = P_\epsilon(\Delta y_i[1], \dots, \Delta y_i[m]) = \frac{1}{(2\pi\sigma_\eta^2)^{m/2}} \exp\left(-\frac{1}{2\sigma_\eta^2} \sum_{k=1}^m \Delta y_i^2[k]\right)$$

entonces, la relación de verosimilitud tiene la forma

$$\begin{aligned} \lambda(y) &= \frac{P_\eta(y - x_1)}{P_\eta(y - x_0)} = \exp\left(-\frac{1}{2\sigma_\eta^2} \sum_{k=1}^m (\Delta y_i^2[k] - \Delta_0^2[k])\right) \\ &= \exp\left(-\frac{1}{N_0} \sum_{k=1}^m ((y[k] - x_1[k])^2 - (y[k] - x_0[k])^2) \Delta t\right) \end{aligned} \quad (1)$$

donde  $N_0 = \sigma_\eta^2/\Delta f = 2\sigma_\eta^2\Delta t$  es la densidad espectral de potencia del ruido.

Teniendo en cuenta (1), el criterio de detección (6.18), toma la forma

$$\ln \lambda(y) = -\frac{1}{N_0} \sum_{k=1}^m ((y[k] - x_1[k])^2 - (y[k] - x_0[k])^2) \Delta t \underset{H_0}{\overset{H_1}{\gtrless}} \ln \gamma$$

asumiendo que el ancho de banda  $\Delta f$  es suficientemente grande, y por lo tanto,  $\Delta t \rightarrow 0$ , entonces la sumatoria anterior se reemplaza por la operación de integración,

$$-\frac{1}{N_0} \int_0^T ((y(t) - x_1(t))^2 - (y(t) - x_0(t))^2) dt \underset{H_0}{\overset{H_1}{\gtrless}} \ln \gamma \quad (2)$$

## Problemas

**Problema 6.1.** Desarrollar el algoritmo de detección, dado el RBG con valor medio cero y matriz de correlación  $\mathbf{R}$ , para el proceso aleatorio descrito en la forma:

$$y(t, \vartheta) = \sum_{k=1}^m \vartheta_k x_k(t) = \vartheta \mathbf{x}(t)$$

donde  $\mathbf{x}(t) = \{x_k(t) : k = 1, \dots, m\}$  es un sistema dado de funciones determinísticas linealmente independientes,  $\vartheta = \{\vartheta_k : k = 1, \dots, m\}$  es el vector de parámetros aleatorios con FDP conocida  $p_k(\vartheta)$ . La hipótesis  $H_0$  consiste en que la trayectoria  $\xi = (x_1, x_2, \dots, x_N)$  corresponde solo al ruido, mientras, la hipótesis alternativa  $H_1$  corresponde a la presencia del proceso aleatorio mas RBG. Hallar la relación de verosimilitud.

## 6.2. Estimación de parámetros en señales aleatorias

### 6.2.1. Estimación en condiciones de ruido aditivo

Sea una trayectoria de un proceso aleatorio,  $y(t) \in \mathfrak{Y}$ , que se observa en el intervalo  $(0, T)$ . Se asume, de forma a priori, que el proceso corresponde a una mezcla aditiva de una señal  $x(\boldsymbol{\vartheta}, t)$  con estructura determinística más una perturbación,  $\xi(t)$ :

$$y(t) = x(\boldsymbol{\vartheta}, t) + \xi(t) \quad (6.19)$$

donde  $\boldsymbol{\vartheta} = (\vartheta_1, \dots, \vartheta_m)$ ,  $\boldsymbol{\vartheta} \in \Theta$  es el vector de parámetros desconocidos de la señal, cuyo respectivo vector de estimación es un funcional de la trayectoria observada:

$$\tilde{\boldsymbol{\vartheta}} = \mathcal{K} \{y(t)\}$$

Por cierto,  $\tilde{\boldsymbol{\vartheta}} \in \Theta$ , en la medida en que usualmente el espacio de estimación coincide con el mismo espacio de los parámetros estimados. El valor medio de la perturbación se asume igual a cero,  $E\{\xi(t)\} = 0$ , y con función conocida de correlación propia,  $R_\xi(t, \tau)$ .

La estimación de los parámetros se puede realizar por cualquiera de los criterios de aproximación, descritos en el numeral §2.3. En el caso particular del empleo del criterio de máxima verosimilitud, dado el modelo (6.19), se obtiene el siguiente logaritmo para el funcional de verosimilitud:

$$\ln \Lambda(y(t)|\boldsymbol{\vartheta}) = \int_0^T v(t, \boldsymbol{\vartheta}) \left( y(t) - \frac{1}{2}x(t, \boldsymbol{\vartheta}) \right) dt \quad (6.20)$$

donde  $v(t, \boldsymbol{\vartheta})$  es la solución de la ecuación integral lineal no homogénea

$$\int_0^T R_\xi(t, \tau) v(\boldsymbol{\vartheta}, \tau) d\tau = x(t, \boldsymbol{\vartheta}), \quad 0 < t \in T \quad (6.21)$$

### 6.2.2. Estimación de máxima verosimilitud

La estimación, en este caso, se obtiene de la solución del respectivo sistema de ecuaciones (2.68), para lo cual inicialmente se determinan las derivadas parciales por cada uno de los parámetros del logaritmo del funcional de la relación de verosimilitud (6.20),

$$\begin{aligned} \frac{\partial}{\partial \vartheta_i} \ln \Lambda(y(t)|\boldsymbol{\vartheta}) &= \int_0^T \frac{\partial}{\partial \vartheta_i} v(t, \boldsymbol{\vartheta}) \left( y(t) - \frac{1}{2}x(t, \boldsymbol{\vartheta}) \right) dt \\ &\quad - \frac{1}{2} \int_0^T v(t, \boldsymbol{\vartheta}) \frac{\partial}{\partial \vartheta_i} x(t, \boldsymbol{\vartheta}) dt, \quad i = 1, \dots, m \end{aligned} \quad (6.22)$$

Sin embargo, de (6.21) se tiene que

$$\begin{aligned}
 \int_0^T v(t, \boldsymbol{\vartheta}) \frac{\partial}{\partial \vartheta_i} x(t, \boldsymbol{\vartheta}) dt &= \int_0^T v(t, \boldsymbol{\vartheta}) \left( \int_0^T R_\xi(t, u) \frac{\partial}{\partial \vartheta_i} v(u, \boldsymbol{\vartheta}) du \right) dt \\
 &= \int_0^T \frac{\partial}{\partial \vartheta_i} v(u, \boldsymbol{\vartheta}) \left( \int_0^T R_\xi(t, u) v(t, \boldsymbol{\vartheta}) dt \right) du \\
 &= \int_0^T \frac{\partial}{\partial \vartheta_i} v(u, \boldsymbol{\vartheta}) x(u, \boldsymbol{\vartheta}) du
 \end{aligned} \tag{6.23}$$

Reemplazando (6.23) en (6.22), entonces, se obtiene

$$\frac{\partial}{\partial \vartheta_i} \ln \Lambda(x(t)|\boldsymbol{\vartheta}) = \int_0^T \frac{\partial}{\partial \vartheta_i} v(u, \boldsymbol{\vartheta}) (y(t) - x(t, \boldsymbol{\vartheta})) dt, \quad i = 1, \dots, m \tag{6.24}$$

A partir de la expresión (6.24), se obtiene directamente el sistema de ecuaciones de máxima verosimilitud:

$$\int_0^T \frac{\partial}{\partial \vartheta_i} v(u, \boldsymbol{\vartheta}) (y(t) - x(t, \boldsymbol{\vartheta})) dt = 0, \quad i = 1, \dots, m \tag{6.25}$$

La solución para el sistema de ecuaciones de (6.25), en función de cada una de las variables  $(\vartheta_1, \dots, \vartheta_m)$ , corresponde a la estimación de máxima verosimilitud para el vector de parámetros  $\boldsymbol{\vartheta}$  de la señal,  $\tilde{\boldsymbol{\vartheta}} = (\tilde{\vartheta}_1, \dots, \tilde{\vartheta}_m)$ . La exigencia (2.69), sobre la consistencia de la estimación, impone una condición adicional por la cual la matriz informativa de Fisher:

$$\begin{aligned}
 I(\boldsymbol{\vartheta}; i, j) &= E \left\{ \frac{\partial}{\partial \vartheta_i} \ln \Lambda(y(t)|\boldsymbol{\vartheta}) \frac{\partial}{\partial \vartheta_j} \ln \Lambda(y(t)|\boldsymbol{\vartheta}) \right\} \\
 &= \int_0^T \int_0^T \frac{\partial}{\partial \vartheta_i} v(\tau_1, \boldsymbol{\vartheta}) \frac{\partial}{\partial \vartheta_j} v(\tau_2, \boldsymbol{\vartheta}) R_\xi(\tau_1, \tau_2) d\tau_1 d\tau_2, \quad i, j = 1, \dots, m
 \end{aligned}$$

debe ser positiva definida para  $\tilde{\boldsymbol{\vartheta}} = \boldsymbol{\vartheta}$ .

Cuando la perturbación  $\xi(t)$  corresponde al RBG, el cual tiene DEP igual a  $N_0/2$ , entonces de (6.25) se tiene que  $v(t, \boldsymbol{\vartheta}) = x(t, \boldsymbol{\vartheta})/N_0$ , con lo cual, el sistema de ecuaciones de máxima verosimilitud se simplifica hasta

$$\int_0^T \frac{\partial}{\partial \vartheta_i} x(t, \boldsymbol{\vartheta}) (y(t) - x(t, \boldsymbol{\vartheta})) dt = 0, \quad i = 1, \dots, m$$

En consecuencia, al ruido blanco Gaussiano le corresponden los siguientes elementos para



la matriz de Fisher:

$$I(\boldsymbol{\vartheta}; i, j) = \frac{1}{N_0} \int_0^T \frac{\partial}{\partial \vartheta_i} x(\tau, \boldsymbol{\vartheta}) \frac{\partial}{\partial \vartheta_j} x(\tau, \boldsymbol{\vartheta}) d\tau, \quad i, j = 1, \dots, m$$

### 6.2.3. Combinación lineal de señales

Sea la relación lineal entre los parámetros de la señal, donde  $x_i(t)$  son funciones conocidas, expresada como:

$$x(t, \boldsymbol{\vartheta}) = \sum_{j=1}^m \vartheta_j x_j(t), \quad (6.26)$$

La estimación conjunta de máxima verosimilitud para  $\vartheta_1, \dots, \vartheta_m$  se obtiene de reemplazar (6.26) en la parte derecha de (6.21), con lo cual se obtiene el sistema de ecuaciones:

$$\int_0^T R_\xi(t, \tau) v_i(\tau) d\tau = x_i(t), \quad i = 1, \dots, m \quad (6.27)$$

La función  $v(t, \boldsymbol{\vartheta})$ , de la que depende el logaritmo del funcional de verosimilitud, es:

$$v(t, \boldsymbol{\vartheta}) = \sum_{j=1}^m \vartheta_j v_j(t)$$

Por lo tanto,

$$\frac{\partial}{\partial \vartheta_i} v(t, \boldsymbol{\vartheta}) = v_i$$

que al reemplazar en (6.25), resulta en el sistema de ecuaciones de máxima verosimilitud,

$$\int_0^T v_i(t) \left( y(t) - \sum_{j=1}^m \vartheta_j x_j(t) \right) dt = 0, \quad i = 1, \dots, m \quad (6.28)$$

$$\text{luego,} \quad \sum_{j=1}^m \vartheta_j \int_0^T v_i(t) x_j(t) dt = \int_0^T v_i(t) y(t) dt, \quad i = 1, \dots, m.$$

Al tomar las notaciones

$$x_{ij}(T) = \int_0^T v_i(t) x_j(t) dt \quad (6.29a)$$

$$y_i(T) = \int_0^T v_i(t) y(t) dt \quad (6.29b)$$

entonces, el sistema lineal de ecuaciones (6.28) se puede escribir en la forma:

$$\sum_{j=1}^m x_{ij}(T) \vartheta_j = y_i(T), \quad i = 1, \dots, m$$

que en forma matricial es

$$\mathbf{X}_{m \times m}(T) \boldsymbol{\vartheta}_{m \times 1} = \mathbf{y}_{1 \times m}(T) \quad (6.30)$$

Asumiendo que  $\int_0^T x_j^2(t) dt < \infty$ ,  $\forall j$ , además que  $R_\xi(t, \tau) > 0$ , entonces, se concluye que existe la matriz inversa  $\mathbf{X}^{-1}(T)$ , razón por la cual la solución de (6.30) conlleva a las siguientes estimaciones de máxima verosimilitud para los parámetros desconocidos:

$$\tilde{\boldsymbol{\vartheta}} = \mathbf{X}^{-1}(T) \mathbf{y}(T) \quad (6.31)$$

La comprobación sobre la consistencia de la estimación, se realiza empleando la parte izquierda de (6.28), de lo cual se encuentran los siguientes elementos informativos de la matriz de Fisher:

$$I(i, j) = E \left\{ \int_0^T v_i(\tau_1) \left( y(\tau_1) - \sum_{k=1}^m \vartheta_k x_k(\tau_1) \right) d\tau_1 \times \right. \\ \left. \int_0^T v_j(\tau_2) \left( y(\tau_2) - \sum_{n=1}^m \vartheta_n x_n(\tau_2) \right) d\tau_2 \right\}, \quad i, j = 1, \dots, m \quad (6.32)$$

De la expresión (6.32), se observa que la matriz informativa de Fisher es positiva definida, debido a que la función de correlación de la perturbación también es positiva definida. Por cierto, los elementos de la matriz informativa, en este caso, no dependen de los parámetros  $\vartheta_1, \dots, \vartheta_m$ . Por lo tanto, la solución (6.31) realmente representa una estimación consistente de máxima verosimilitud para el vector de parámetros  $\boldsymbol{\vartheta}$ . Además, las partes derechas de las expresiones (6.29a) y (6.32) coinciden, con lo cual la matriz informativa de Fisher  $I(T)$  es igual a la matriz  $\mathbf{X}(T)$ , entonces, la expresión (6.31) se puede representar en la forma

$$\tilde{\boldsymbol{\vartheta}} = I^{-1}(T) \mathbf{y}(T)$$

En el caso particular del RBG, al cual le corresponde una DEP igual a  $N_0$ , de (6.27) se obtiene que  $v_i(t) = x_i(t)/N_0$ ,  $i = 1, \dots, m$ , donde

$$x_{ij}(T) = \frac{1}{N_0} \int_0^T x_i(t) x_j(t) dt = I(i, j; T), \quad i, j = 1, \dots, m$$

$$y_i(T) = \frac{1}{N_0} \int_0^T x_i(t) y(t) dt, \quad i = 1, \dots, m$$

Cabe anotar que, en general, la estimación (6.31) no es centrada, en la medida en que

$$E\{\tilde{\boldsymbol{\theta}}\} = \mathbf{X}^{-1}(T)E\{\mathbf{y}(T)\} \quad (6.33)$$

Sin embargo,

$$E\{y_i(T)\} = \sum_{j=1}^m \vartheta_j \int_0^T x_j(t)v_i(t)dt = \sum_{j=1}^m \vartheta_j x_{ij}(T)$$

entonces

$$E\{\mathbf{X}(T)\} = \mathbf{y}(T)\boldsymbol{\vartheta} \quad (6.34)$$

Reemplazando (6.34) en (6.33), finalmente se obtiene que  $E\{\tilde{\boldsymbol{\theta}}\} = \boldsymbol{\vartheta}$  lo que demuestra la ausencia de sesgo en la estimación (6.31).

**Ejemplo 6.2.** Hallar la estimación de máxima verosimilitud para una señal con amplitud desconocida,  $\mathbf{ax}(t)$ , que está inmersa en ruido blanco Gaussiano.

La señal en análisis se considera un caso particular de (6.26), cuando  $m = 1$  y  $\vartheta_1 = a$ . Luego, la estimación para el respectivo modelo lineal de señal en (6.31) toma la forma:

$$\tilde{a} = \int_0^T v(t)y(t)dt \Big/ \int_0^T v(t)x(t)dt \quad (6.35)$$

donde

$$y(T) = \int_0^T v(t)y(t)dt, \quad x(T) = \int_0^T v(t)x(t)dt$$

siendo  $v(t)$  la solución de la ecuación lineal integral dada en (6.27),

$$\int_0^T R_{\xi}(t, \tau)v(\tau)d\tau = x(t), \quad 0 \leq t \leq T$$

Como antes se indicó, la estimación (6.35) de máxima verosimilitud para un parámetro del modelo lineal de la señal no tiene sesgo y es efectiva, esto es

$$E\{\tilde{a}\} = a$$

$$\mu_2\{\tilde{a}\} = \mathbf{X}_T^{-1} = I^{-1}(T) = \left( \int_0^T v(t)x(t)dt \right)^{-1}$$

Al asumir el modelo del ruido blanco Gaussiano, entonces,  $v(t) = x(t)/N_0$ , y de (6.35) se

tiene, entonces, que

$$\tilde{a} = \int_0^T x(t)y(t)dt \Big/ \int_0^T x^2(t)dt$$

La implementación del algoritmo de máxima verosimilitud para la estimación de la amplitud de la señal, cuando se tiene RBG, exige el cálculo de la integral normalizada de correlación (6.35). Esta operación se realiza mediante la filtración lineal con respuesta a impulso [12]:

$$h(\tau) = \begin{cases} x(T - \tau) \Big/ \int_0^T x^2(t)dt, & 0 \leq \tau \leq T \\ 0, & \tau < 0, \tau > T \end{cases}$$

### Problemas

**Problema 6.2.** Dado el modelo de combinación lineal de señales (6.26),

- Demostrar que la estimación vectorial de máxima verosimilitud (6.31) no tiene sesgo.
- Hallar la matriz de correlación de la estimación de máximo verosimilitud  $E\{(\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta})(\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta})^T\}$  y demostrar la efectividad de la estimación (6.31).

**Problema 6.3.** Asumir que se tiene una trayectoria  $y(t)$  de la suma de señal y RBG (6.31), observada en el intervalo de tiempo  $(0, T)$ , la cual es discretizada para obtener la correspondiente sucesión  $y = \{y(t_i) : i = 1, \dots, n\}$ ,  $t_i \in T$ :

- Hallar la correspondiente estimación de máximo verosimilitud.
- Demostrar que la estimación obtenida es insesgada y hallar su matriz de correlación.

**Problema 6.4.** Al igual que en el problema anterior, asumir la discretización de la trayectoria de la señal  $ax(t)$  del ejemplo 6.2, para la estimación de la amplitud desconocida  $a$ . Hallar la estimación de máximo verosimilitud.

### 6.3. Filtración óptima lineal por mínimos cuadrados

En este caso, la filtración optimiza la forma de representación de los dispositivos lineales de proceso, como observa en la Figura 6.2, en la cual las observaciones iniciales  $x$  pertenecientes al proceso  $\xi$  interactúan con la perturbación  $\eta$ , en forma general se analiza la combinación lineal; acción que se representa mediante el funcional de transformación  $\mathcal{K}_1\{\xi, \eta\}$ , cuya señal de salida pasa por el filtro lineal, representado por la respuesta a impulso  $h(t, s)$  y que se escoge, de tal manera que la salida del filtro en el momento  $t$  corresponda al mínimo del error cuadrático medio (2.64) de la estimación  $\theta \subset x$ , que se contiene en la señal  $\xi$ . El mismo parámetro en estimación se puede describir por la transformación  $\mathcal{K}_0\{\xi\}$ .

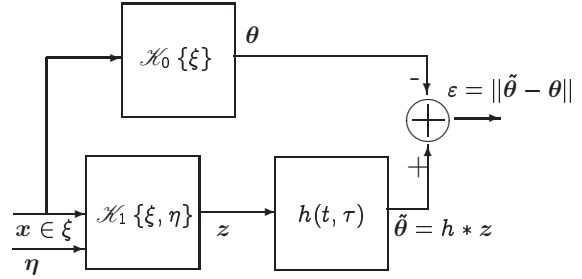


Figura 6.2. Filtración de señales

Cuando se analiza la transmisión de señales, durante la cual es importante asegurar el mínimo de distorsión, el operador de transformación  $\mathcal{K}_0$  describe un canal ideal, y por lo tanto, se hace necesario sintetizar un filtro con respuesta a impulso, conectado en serie al canal real con operador de transformación  $\mathcal{K}_1$ , de tal manera que su salida corresponda a alguna transformación tan cercana como se pueda a  $\mathcal{K}_0$ .

#### 6.3.1. Optimización de la respuesta a impulso

Sea  $z(t)$  una trayectoria para la combinación lineal en forma de proceso aleatorio aditivo:  $z(t) = x(t) + \eta(t)$ , tal que  $\|x\|, \|\eta\| < \infty$ . En calidad de estimación  $\tilde{\theta}(t)$  se toma el valor filtrado de la observación:

$$\tilde{\theta}(t) = \int_{-\infty}^{\infty} h(t, \tau) z(\tau) d\tau \quad (6.36)$$

siendo  $h(t, \tau)$  la respuesta a impulso del dispositivo lineal, que por ahora se asume sin restricciones de causalidad en su implementación práctica. No obstante, el filtro se considera que cumple con la condición de estabilidad.

La estimación (6.36) es ante todo sesgada, en la medida en que

$$m_{1\tilde{\theta}}(t) = \int_{-\infty}^{\infty} h(t, \tau) (m_{1x}(\tau) + m_{1\eta}(\tau)) d\tau \neq m_{1x}(t)$$

donde  $m_{1\eta}(t)$  y  $m_{1x}(t)$  son las respectivas medias de la señal útil y de la perturbación.

El valor cuadrático medio del error, determinado como  $\varepsilon = (\theta - \tilde{\theta})^2$ , se obtiene mediante

su promedio de tiempo, que en caso emplear la estimación (6.36), es igual a:

$$\varepsilon^2(t) = E \left\{ \left( \theta(t) - \tilde{\theta}(t) \right)^2 \right\} \quad (6.37)$$

Al tomar el error, (6.37), como criterio de calidad se puede encontrar la respuesta a impulso que brinde su mínimo valor con respecto a todos los demás posibles sistemas lineales. En particular, al reemplazar (6.36) en (6.37), se obtiene:

$$\begin{aligned} \varepsilon^2(t) &= E \left\{ \theta^2(t) - 2 \int_{-\infty}^{\infty} h(t, \tau) z(\tau) \theta(t) d\tau \right. \\ &\quad \left. + \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h(t, \lambda_1) h(t, \lambda_2) z(\lambda_1) z(\lambda_2) d\lambda_1 d\lambda_2 \right\} \\ &= E \left\{ \theta^2(t) \right\} - 2 \int_{-\infty}^{\infty} h(t, \tau) E \{ z(\tau) \theta(t) \} d\tau + \\ &\quad + \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h(t, \lambda_1) h(t, \lambda_2) E \{ z(\lambda_1) z(\lambda_2) \} d\lambda_1 d\lambda_2 \end{aligned} \quad (6.38)$$

Sean conocidas las funciones de correlación propia,  $R_i(t_1, t_2)$ ,  $i = \{z, \theta\}$ , además, de la respectiva función de correlación mutua  $R_{z\theta}(t_1, t_2)$  entre los procesos  $z(t)$  y  $\theta(t)$ .

Por cuanto,

$$\begin{aligned} E \{ z(\lambda_1) z(\lambda_2) \} &= R_z(\lambda_1, \lambda_2) \\ &= R_x(\lambda_1, \lambda_2) + R_\xi(\lambda_1, \lambda_2) + R_{\xi x}(\lambda_1, \lambda_2) + R_{x\xi}(\lambda_1, \lambda_2) \end{aligned} \quad (6.39a)$$

$$E \{ z(\tau) \theta(t) \} = R_{z\theta}(\lambda_1, \lambda_2) = R_\theta(\tau, t) + R_{z\theta}(\tau, t) \quad (6.39b)$$

entonces, reemplazando (6.39a) y (6.39b) en (6.38) se obtiene que

$$\begin{aligned} \varepsilon^2(t) &= R_\theta(t, t) - 2 \int_{-\infty}^{\infty} h(t, \tau) R_{z\theta}(\tau, t) d\tau \\ &\quad + \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h(t, \lambda_1) h(t, \lambda_2) R_z(\lambda_1, \lambda_2) d\lambda_1 d\lambda_2 \end{aligned} \quad (6.40)$$

La expresión (6.40) muestra que el error cuadrático medio de la estimación lineal  $\tilde{\theta}(t)$  depende, tanto de las funciones de correlación propia  $R_i(t_1, t_2)$ ,  $i = \{z, \theta\}$ , como de la función de correlación mutua  $R_{z\theta}(t_1, t_2)$  de los procesos  $z(t)$  y  $\theta(t)$ , pero de ninguna manera de la estructura más fina de estos procesos [12]. En general, se puede demostrar que, dadas todas las anteriores funciones de correlación, la mejor estimación lineal en el sentido del error cuadrático medio corresponde al filtro con respuesta a impulso que cumpla

la ecuación integral [43]:

$$R_{z\theta}(\tau, t) = \int_{-\infty}^{\infty} h_{opt}(t, \lambda) R_z(\tau, \lambda) d\lambda \quad (6.41)$$

El reemplazo de (6.41) en (6.40), resulta en

$$\begin{aligned} \varepsilon^2(t) = R_\theta(t, t) - 2 \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h(t, \tau) h_{opt}(t, \lambda) R_z(\tau, \lambda) d\tau d\lambda \\ + \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h(t, \lambda_1) h(t, \lambda_2) R_z(\lambda_1, \lambda_2) d\lambda_1 d\lambda_2 \end{aligned}$$

luego

$$\begin{aligned} \varepsilon^2(t) = R_\theta(t, t) - \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h_{opt}(t, \tau) h_{opt}(t, \lambda) R_z(\tau, \lambda) d\tau d\lambda \\ + \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} R_z(\lambda_1, \lambda_2) \{h(t, \lambda_1) - h_{opt}(t, \lambda_1)\} \{h(t, \lambda_2) - h_{opt}(t, \lambda_2)\} d\lambda_1 d\lambda_2 \end{aligned} \quad (6.42)$$

Por cuanto, sólo el último término de (6.42) contiene la función desconocida  $h(t, \lambda)$ , la cual debe ser definida positiva [12], entonces el menor valor de  $\overline{\varepsilon^2(t)}$  ocurre cuando el último término es 0, esto es, cuando el filtro tiene respuesta a impulso  $h(t, \lambda) \equiv h_{opt}(t, \lambda)$ . En este caso, el valor mínimo de error será:

$$\varepsilon_{\min}^2(t) = R_\theta(t, t) - \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h_{opt}(t, \tau) h_{opt}(t, \lambda) R_z(\tau, \lambda) d\tau d\lambda$$

Teniendo en cuenta (6.41), entonces

$$\varepsilon_{\min}^2(t) = R_\theta(t, t) - \int_{-\infty}^{\infty} h_{opt}(t, \lambda) R_{z\theta}(\lambda, t) d\lambda$$

De otra parte, la expresión (5.8) relaciona las funciones de correlación propia a la entrada y salida de un sistema lineal, con lo cual, la última expresión toma la forma definitiva,

$$\varepsilon_{\min}^2(t) = R_\theta(t, t) - R_{\hat{\theta}}(t, t) \quad (6.43)$$

la cual implica, en primer orden, que el mínimo error cuadrático medio (2.64) corresponde a la diferencia de los valores cuadráticos medios del proceso a estimar y de su estimación. En segundo lugar, se tiene que  $R_{\hat{\theta}}(t, t) \leq R_\theta(t, t)$ .

Si se asume la estacionariedad de los procesos  $\theta(t)$  y  $\eta(t)$ , por lo menos en el sentido amplio, además el filtro se supone con parámetros en (1.38) invariables en el tiempo, entonces la ecuación integral (6.41) toma la forma denominada *ecuación de Wiener-Hopf*:

$$R_{z\theta}(\tau) = \int_{-\infty}^{\infty} h_{opt}(\lambda) R_z(\tau - \lambda) d\lambda \quad (6.44)$$

La invariabilidad de los momentos de aleatoriedad de los procesos estacionarios, supone la invariabilidad en el valor del error cuadrático medio, cuyo valor mínimo (6.43), es:

$$\begin{aligned} \varepsilon_{\min}^2 &= R_{\theta}(0) - \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h_{opt}(\lambda_1) h_{opt}(\lambda_2) R_z(\lambda_1 - \lambda_2) d\lambda_1 d\lambda_2 \\ &= R_{\theta}(0) - \int_{-\infty}^{\infty} h_{opt}(\lambda) R_{z\theta}(\lambda) d\lambda \end{aligned} \quad (6.45)$$

con lo que el error cuadrático medio es igual a

$$\varepsilon_{\min}^2 = R_{\theta}(0) - R_{\tilde{\theta}}(0) = m_{2\theta} - m_{2\tilde{\theta}} \quad (6.46)$$

esto es, se determina por la diferencia de los valores medios de potencia, tanto del proceso a estimar, como de la propia estimación.

En concordancia con (3.24), los valores medios de potencia se pueden expresar a través de las respectivas densidades espectrales de potencia,

$$\varepsilon_{\min}^2 = \frac{1}{2\pi} \int_0^{\infty} \left( S_{\theta}(\omega) - S_z(\omega) |H_{opt}(\omega)|^2 \right) d\omega \quad (6.47)$$

siendo  $H_{opt}(\omega)$  la función de transferencia del filtro óptimo lineal, además

$$S_z(\omega) = S_{\theta}(\omega) + S_{\eta}(\omega) + S_{\theta\eta}(\omega) + S_{\eta\theta}(\omega) \quad (6.48)$$

donde  $S_{\theta}(\omega)$ ,  $S_{\eta}(\omega)$ ,  $S_{\theta\eta}(\omega)$  y  $S_{\eta\theta}(\omega)$  son las densidades espectrales de potencia propias y mutuas de los procesos  $\theta(t)$  y  $\eta(t)$ , de manera correspondiente.

De esta manera, la respuesta a impulso óptima de un sistema lineal, en caso de asumir la estacionariedad de los procesos, consiste en la solución de la ecuación integral (6.44), sin tener en cuenta la condición de causalidad, mediante la TF en la forma,

$$S_{z\theta}(\omega) = S_{\theta}(\omega) + S_{z\theta}(\omega) = H_{opt}(\omega) S_z(\omega)$$

de donde, teniendo en cuenta (6.48), se tiene

$$H_{opt}(\omega) = \frac{S_{z\theta}(\omega)}{S_z(\omega)} = \frac{S_{\theta}(\omega) + S_{z\theta}(\omega)}{S_{\theta}(\omega) + S_{\eta}(\omega) + S_{\theta\eta}(\omega) + S_{\eta\theta}(\omega)} \quad (6.49)$$



Reemplazando (6.49) en (6.47) se obtiene el valor del mínimo error cuadrático medio,

$$\varepsilon_{\min}^2 = \frac{1}{2\pi} \int_0^\infty \frac{S_\theta(\omega) S_z(\omega) - |S_{z\theta}(\omega)|^2}{S_z(\omega)} d\omega \quad (6.50)$$

Si se considera la correlación nula entre las señales  $\theta(t)$  y  $\eta(t)$ , esto es,  $R_{\theta\eta}(\tau) = 0$ , entonces, sus DEP mutuas también son nulas,  $S_{\theta\eta}(\omega) = S_{\eta\theta}(\omega) = 0$ , por lo que la función de transferencia óptima de un sistema lineal tiene la forma,

$$H_{opt}(\omega) = \frac{S_\theta(\omega)}{S_\theta(\omega) + S_\eta(\omega)}$$

De (6.50), y reemplazando  $S_{z\theta}(\omega) = S_\theta(\omega)$ , además  $S_z(\omega) = S_\theta(\omega) + S_\eta(\omega)$ , se obtiene para las señales de correlación nula,

$$\varepsilon_{\min}^2 = \frac{1}{2\pi} \int_0^\infty \frac{S_\theta(\omega) S_\eta(\omega)}{S_\theta(\omega) + S_\eta(\omega)} d\omega \quad (6.51)$$

El error cuadrático medio (6.51) se puede hacer igual a cero, cuando las densidades espectrales de potencia de las señales  $\theta(t)$  y  $\eta(t)$  no se traslapan, esto es, cuando para todo valor de  $\omega$  se cumple que  $S_\eta(\omega) = S_\theta(\omega) = 0$ , lo cual a su vez implica que por lo menos uno de los dos espectros debe tener ancho de banda finito. En caso contrario, siempre ocurrirá un error.

En síntesis, la filtración de una señal estacionaria a partir de su combinación lineal con otro proceso estacionario, de tal manera que el error cuadrático medio sea 0, corresponde al caso en que las respectivas DEP de ambos procesos tienen rangos de frecuencia en los cuales no hay aporte energético. En este sentido, si se analiza la filtración del proceso  $\theta(t)$  dado sobre un fondo de ruido blanco Gaussiano, entonces de (6.51) se obtiene

$$\varepsilon_{\min}^2 = \frac{N_0}{2\pi} \int_0^\infty \frac{S_\theta(\omega)}{S_\theta(\omega) + N_0} d\omega$$

que implica, que siempre se tiene en condiciones reales un valor diferente de 0 para el error cuadrático medio.

### 6.3.2. Condición de realización física

Un filtro lineal con respuesta a impulso óptima (6.41) no es físicamente realizable, al no cumplir la condición de causalidad, por cuanto  $h_{opt}(t, \lambda) \neq 0$ , para  $t < \lambda$ . Al observar en el numeral §6.3.1, la operación de filtración se realiza después de analizada la combinación lineal de las señales  $\theta(t)$  y  $\eta(t)$  sobre todo el intervalo de tiempo  $t \in (-\infty, \infty)$ , lo cual implica que la filtración óptima realiza la estimación del valor del proceso en un momento dado del tiempo con retardo infinito.

La condición de realización física implica que la filtración debe realizarse sobre la observación  $z(t)$  determinada hasta el momento de tiempo, en el cual se realiza la estimación,

$$h(t, \tau) = 0, \quad \forall \tau < 0 \quad (6.52)$$

Reemplazando la condición de causalidad (6.52) en (6.36) se obtiene la estimación de  $\theta(t)$  con un filtro físicamente realizable,

$$\tilde{\theta}(t) = \int_{-\infty}^t h(t, \tau) z(\tau) d\tau$$

Frecuentemente, la estimación se realiza sobre intervalos cerrados y finitos de análisis,  $(t - T, t)$ , por lo que para la anterior integral se ajustan los límites en la siguiente forma,

$$\tilde{\theta}(t) = \int_{t-T}^t h(t, \tau) z(\tau) d\tau = \int_0^T h(t, t - \lambda) z(t - \lambda) d\lambda \quad (6.53)$$

El reemplazo de (6.53) en (6.37), y al efectuar las mismas transformaciones de (6.38) se obtiene la correspondiente expresión del error cuadrático medio del filtro óptimo que se ajusta a la condición de realización física, asumiendo un intervalo finito de estimación:

$$\begin{aligned} \varepsilon^2(t) = & R_{\theta}(t, t) - 2 \int_0^T h(t, t - \tau) R_{z\theta}(t - \tau, t) d\tau + \\ & + \int_0^T \int_0^T h(t, t - \lambda_1) h(t, t - \lambda_2) R_z(t - \lambda_1, t - \lambda_2) d\lambda_1 d\lambda_2 \end{aligned} \quad (6.54)$$

El filtro lineal óptimo para la definición del error acotado (6.54), implica que su respuesta impulso causal debe cumplir la ecuación integral,

$$R_{z\theta}(t - \tau, t) = \int_T^t h_{opt}(t, t - \lambda) R_z(t - \tau, t - \lambda) d\lambda, \quad \tau \in T \quad (6.55)$$

con lo cual el valor mínimo del error cuadrático medio del filtro óptimo físicamente implementable es igual a

$$\varepsilon_{\min}^2(t) = R_{\theta}(t, t) - \int_0^T h_{opt}(t, t - \lambda) R_{z\theta}(t - \lambda, t) d\lambda$$

Al asumir la estacionariedad de los procesos  $\theta(t)$  y  $\eta(t)$ , además suponiendo la invariabilidad de sus parámetros en el tiempo, entonces, las anteriores dos relaciones toman la

siguiente forma definitiva:

$$R_{z\theta}(\tau) = \int_T h_{opt}(\lambda) R_z(\tau - \lambda) d\lambda, \tau \in T \quad (6.56a)$$

$$\varepsilon_{\min}^2 = R_\theta(0) - \int_T h_{opt}(\lambda) R_{z\theta}(\lambda) d\lambda \quad (6.56b)$$

**Ejemplo 6.3.** Sea la señal  $\theta = as(t)$ , siendo  $s(t)$  una función conocida  $a$  y la amplitud aleatoria. Hallar la respuesta del filtro lineal que minimize el error cuadrático medio de la estimación  $\tilde{\theta} = \tilde{a}s(t)$  hecha de la trayectoria  $\xi(t) + \eta(t) \in (t-T, t)$ , siendo  $\eta(t)$  un proceso aleatorio independiente de  $a$ , con  $\overline{\eta(t)} = 0$  y función de correlación  $R_\eta(t_1, t_2)$ .

La solución de la ecuación integral (6.55), reemplazando  $R_{z\theta}(t_1, t_2)$  y  $R_\theta(t_1, t_2)$  por:

$$\begin{aligned} R_{z\theta}(t - \tau, t) &= E\{a^2\} s(t - \tau) s(t) \\ R_z(t - \tau, t - \lambda) &= E\{a^2\} s(t - \tau) s(t - \lambda) + R_\eta(t - \tau, t - \lambda) \end{aligned}$$

con lo cual, para  $\tau \in T$  se obtiene que

$$E\{a^2\} s(t - \tau) s(t) = \int_0^T h_{opt}(t, t - \lambda) (E\{a^2\} s(t - \tau) s(t - \lambda) - R_\eta(t - \tau, t - \lambda)) d\lambda,$$

La solución de la anterior ecuación integral tiene la forma,

$$h_{opt}(t, \tau) = ks(t)v(t)$$

siendo  $v(t)$  la solución de la ecuación integral [44],

$$\int_{t-T}^t v(\lambda_1) R_\eta(\lambda_2, \lambda_1) d\lambda_1 = s(\lambda_2), \lambda_2 \in (t-T, t)$$

Al reemplazar las anteriores expresiones en la solución buscada de (6.55) se obtiene,

$$E\{a^2\} = k \left( E\{a^2\} \int_0^T v(t - \lambda) s(t - \lambda) d\lambda + 1 \right)$$

de donde se encuentra el siguiente valor de  $k$ :

$$k = E\{a^2\} \left( E\{a^2\} \int_0^T v(t - \lambda) s(t - \lambda) d\lambda + 1 \right)^{-1}$$

El filtro óptimo, en este caso de estimación, tiene respuesta a impulso,

$$h_{opt}(t, \tau) = \frac{E\{a^2\} s(t) v(t)}{1 + E\{a^2\} \int_{t-T}^t v(\lambda) s(\lambda) d\lambda}$$

A partir de la trayectoria  $z(t)$ , que implica la adición del proceso  $\eta(t)$ , observada en el intervalo de tiempo  $(t-T, t)$ , la estimación de la señal en el momento  $t$  tiene la forma:

$$\tilde{a}s(t) = \int_0^T h_{opt}(t, t-\lambda) z(t-\lambda) d\lambda = \frac{E\{a^2\} s(t) \int_{t-T}^t z(\lambda) v(\lambda) d\lambda}{1 + E\{a^2\} \int_{t-T}^t v(\lambda) s(\lambda) d\lambda}$$

por lo cual, la estimación de la misma amplitud de la señal es igual a

$$\tilde{a} = \frac{E\{a^2\} \int_{t-T}^t z(\lambda) v(\lambda) d\lambda}{1 + E\{a^2\} \int_{t-T}^t v(\lambda) s(\lambda) d\lambda}$$

que resulta ser sesgada, como se observa de su valor promedio,

$$E\{\tilde{a}\} = E\{a\} \frac{E\{a^2\} \int_{t-T}^t s(\lambda) v(\lambda) d\lambda}{1 + E\{a^2\} \int_{t-T}^t v(\lambda) s(\lambda) d\lambda}$$

Aunque asintóticamente, cuando  $E\{a^2\} \rightarrow \infty$ , esta puede ser considerada como no sesgada. El respectivo error cuadrático medio de la estimación  $\tilde{a}$  está dado por la expresión:

$$\begin{aligned} \varepsilon_{\min}^2(t) &= E\{a^2\} s^2(t) - k E\{a^2\} \int_0^T s^2(t) s(t-\tau) v(t-\tau) d\tau \\ &= E\{a^2\} s^2(t) \left( 1 - k \int_{t-T}^t v(\lambda) s(\lambda) d\lambda \right) \end{aligned}$$

que al reemplazar el valor de la constante  $k$  resulta en

$$\varepsilon_{\min}^2(t) = E\{(\tilde{a} - a)^2 s^2(t)\} = \frac{E\{a^2\} s^2(t)}{1 + E\{a^2\} \int_{t-T}^t v(\lambda) s(\lambda) d\lambda}$$

En el caso particular, cuando la filtración del parámetro aleatorio se considera inmersa en ruido blanco aditivo con densidad espectral de potencia  $N_0/2$  y función de correlación

$$R_\eta(t, \lambda) = \frac{N_0}{2} \delta(t - \lambda), v(\tau) = \frac{2}{N_0} s(\tau)$$

con lo cual, las correspondientes respuesta a impulso óptima y valor de error cuadrático medio tiene la forma:

$$h_{opt}(t, \tau) = \frac{\frac{2E\{a^2\}}{N_0} s(t) s(\tau)}{1 + \frac{2E\{a^2\}}{N_0} \int_{t-T}^t s^2(\lambda) d\lambda}$$

$$\varepsilon_{\min}^2(t) = \frac{E\{a^2\} s^2(t)}{1 + \frac{2E\{a^2\}}{N_0} \int_{t-T}^t s^2(\lambda) d\lambda}$$

### 6.3.3. Filtros acoplados

En el numeral §6.3, se analiza la calidad de filtración mediante el criterio del error cuadrático medio, que se justifica cuando en el proceso de la señal tiene sentido el análisis de la dependencia del error en el tiempo. En otras tareas, es más importante establecer simplemente la presencia o ausencia de la señal inmersa en la perturbación, lo que puede ser resuelto mediante la filtración que brinde la máxima relación señal/ruido, sin importar cuánto se afecte la señal útil.

Sea  $z(t) = x(t) + \eta(t)$ , la combinación lineal de la señal útil y la perturbación aleatoria  $\eta(t)$ , que se supone estacionaria (en el sentido amplio) con valor medio  $\overline{\eta(t)} = 0$  y función conocida de correlación propia  $R_\eta(\tau)$ . La estimación de la señal en el intervalo de observación  $T$  corresponde al valor

$$\tilde{x}(t) = \int_0^T h(\tau) z(t - \tau) d\tau \quad (6.57)$$

donde  $h(\tau)$  es la función respuesta a impulso que debe ser determinada, asumiendo la linealidad del sistema y la invariabilidad de sus parámetros de definición en el tiempo. La relación anterior se puede escribir de la forma:

$$\tilde{x}(t) = x_1(t) + v(t) \quad (6.58)$$

donde

$$x_1(t) = \int_0^T h(\tau) x(t - \tau) d\tau, \quad v(t) = \int_0^T h(\tau) \eta(t - \tau) d\tau$$

La relación por potencia de señal/ruido se determina como la relación del valor cuadrático medio la señal de salida  $\overline{x_1^2(t)}$ , estimada en un momento determinado del tiempo  $t_0$ , sobre

la varianza del ruido a la salida del filtro  $\sigma_v^2$ , así,

$$S/N = \frac{\overline{x_1^2(t)}}{\sigma_v^2} \quad (6.59)$$

El filtro lineal óptimo, en el sentido del máximo valor de la relación  $S/N \triangleq \mu_{\max}$ , se denomina *filtro acoplado*.

En cualquier sistema lineal con respuesta a impulso  $h(\tau)$ , teniendo en cuenta (6.59), se cumple la siguiente desigualdad:

$$\begin{aligned} \mu_{\max} \sigma_v^2 - x_1^2(t_0) &= \mu_{\max} \int_0^T \int_0^T h(\lambda_1) h(\lambda_2) R_\eta(\lambda_2 - \lambda_1) d\lambda_2 d\lambda_1 - \\ &\quad - \left( \int_0^T h(\tau) x(t_0 - \tau) d\tau \right)^2 \geq 0 \end{aligned} \quad (6.60)$$

de tal manera, que el símbolo de la igualdad, que corresponde al caso del filtro acoplado, ocurre para la función  $h(\tau) = h_{ac}(\tau, t_0)$ , que cumpla con la ecuación integral:

$$\int_0^T h_{ac}(\tau; t_0) R_\eta(t - \tau) d\tau = k(t_0) x(t_0 - t), \quad t \in [0, T] \quad (6.61)$$

siendo  $k(t_0)$  una constante, que para el momento  $t_0$  es igual a

$$k(t_0) = \frac{\int_0^T h_{ac}(\tau; t_0) x(t_0 - \tau) d\tau}{\mu_{\max}} = \frac{x_{1c}(t_0)}{\mu_{\max}} = \frac{\sigma_{vc}^2}{x_{1c}(t_0)}$$

donde la señal  $x_{1c}$  y la varianza  $\sigma_{vc}^2$  están dados a la salida del filtro acoplado. Cabe anotar, que si existe la función  $h_{ac}(\tau, t_0)$  como solución de (6.61), entonces esta misma ecuación se cumple para  $c h_{ac}(\tau, t_0)$ , siendo  $c$  una constante (por cierto, el valor  $\mu_{\max}$  no cambia), con lo cual la respuesta a impulso del filtro acoplado puede ser determinada hasta el valor de su factor de escala.

Una solución directa se puede hallar al asumir que la perturbación corresponde a ruido blanco Gaussiano con función de correlación propia  $R_{eta}(t - \tau) = N_0 \delta(t - \tau)$ . Al despejar la función delta dentro de la integral, mediante la propiedad de selectividad, la respuesta a impulso del filtro acoplado toma la forma,

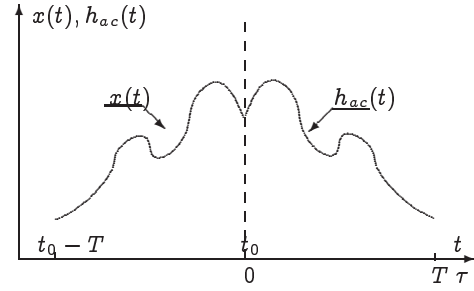
$$h_{ac}(t; t_0) = \frac{k(t_0)}{N_0} x(t_0 - t), \quad t \in [0, T] \quad (6.62)$$

El análisis de la estructura de  $h_{ac}(t; t_0)$  en (6.62), muestra que la respuesta a impulso del filtro acoplado, obtenida en este caso, es directamente proporcional a la forma imagen

con respecto al eje vertical, que pasa por el valor de tiempo  $t = t_0$ , de la señal útil  $x(t)$  en el intervalo  $(t_0 - T, t_0)$ , con el consecuente traslado del origen de las coordenadas hasta el punto  $t = t_0$ . En el caso simple cuando  $t_0 = T$ , la respuesta a impulso del filtro acoplado corresponde a la representación espejo de la señal con respecto al eje vertical, que divide en dos el intervalo de análisis  $(0, T)$ , como se observa en la Figura 6.3.

El máximo valor de la relación señal/ruido, a partir de (6.60), es igual a

$$\begin{aligned}\mu_{\max} &= \frac{\left( \int_0^T x^2(t_0 - \tau) d\tau \right)^2}{N_0 \int_0^T x^2(t_0 - \tau) d\tau} \\ &= \frac{1}{N_0} \int_0^T x^2(t_0 - \tau) d\tau = \frac{\mathcal{E}_x}{N_0}\end{aligned}$$



**Figura 6.3.** Respuesta a impulso de un filtro acoplado

esto es, el valor corresponde a la relación de energía de la señal sobre la DEP del ruido  $N_0$ , en el intervalo  $(t_0 - T, t_0)$ .

La función de transferencia  $H_{ac}(\omega; t_0)$  del filtro acoplado con respuesta a impulso (6.62), se determina como

$$\begin{aligned}H_{ac}(\omega; t_0) &= \int_{-\infty}^{\infty} h_{ac}(t; t_0) e^{-j\omega t} dt = \frac{k(t_0)}{N_0} \int_0^T x(t_0 - t) e^{-j\omega t} dt \\ &= \frac{k(t_0)}{N_0} e^{-j\omega t_0} \int_{t_0 - T}^{t_0} x(t) e^{j\omega t} dt\end{aligned}$$

lo que puede ser interpretado como

$$H_{ac}(\omega; t_0) = \frac{k(t_0)}{N_0} e^{-j\omega t_0} X^*(\omega; t_0, T)$$

Por lo anterior, durante la separación de una señal de su mezcla aditiva con el ruido blanco, la respuesta a impulso del filtro acoplado es proporcional al espectro conjugado  $X^*(\omega; t_0, T)$  de la señal útil y del truncado en el intervalo  $(t_0 - T, t_0)$ .

En forma general, se puede obtener la expresión de la respuesta a impulso para el caso de cualquier DEP de ruido dado  $\mathcal{S}_\eta(\omega)$ , pero sin tener en cuenta la condición para la realización física del filtro, de manera particular, al suponer que se tiene la observación

$z(t)$  para todos los valores de  $t \in \mathbb{R}$ , generalizando los límites de la integral (6.57):

$$\tilde{x}(t) = \int_{-\infty}^{\infty} h(\tau; t_0) z(t - \tau) d\tau$$

En este caso, la ecuación (6.60), para todos los valores de  $t$ , toma la siguiente forma:

$$\int_{-\infty}^{\infty} h_{ac}(\tau; t_0) R_{\eta}(t - \tau) d\tau = k(t_0) x(t_0 - t)$$

que al hallar la transformada de Fourier se convierte en:

$$H_{ac}(\omega; t_0) S_{\eta}(\omega) = k(t_0) X^*(\omega) e^{j\omega t_0}$$

con lo cual, la respuesta a impulso del filtro acoplado generalizado tiene la forma:

$$H_{ac}(\omega; t_0) = k(t_0) \frac{X^*(\omega)}{S_{\eta}(\omega)} e^{-j\omega t_0} \quad (6.63)$$

El máximo valor de la relación señal/ruido, en concordancia con (6.60) se calcula como:

$$\begin{aligned} \mu_{\max} &= \frac{\left( \int_{-\infty}^{\infty} h_{ac}(\tau, t_0) x(t_0 - \tau) d\tau \right)^2}{\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h_{ac}(\lambda_1; t_0) h_{ac}(\lambda_2; t_0) R_{\eta}(\lambda_2 - \lambda_1) d\lambda_1 d\lambda_2} \\ &= \frac{\left( \frac{1}{2\pi} \int_{-\infty}^{\infty} H_{ac}(\omega; t_0) X(\omega) e^{j\omega t} d\omega \right)^2}{\frac{1}{2\pi} \int_{-\infty}^{\infty} S_{\eta}(\omega) |H_{ac}(\omega; t_0)|^2 d\omega} \end{aligned}$$

valor en el cual, al sustituir  $H_{ac}(\omega; t_0)$  por (6.63), se obtiene la expresión final

$$\mu_{\max} = \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{|X(\omega)|^2}{S_{\eta}(\omega)} d\omega \quad (6.64)$$

Cuando la perturbación corresponde al ruido blanco Gaussiano, de la expresión (6.64) resulta entonces el siguiente valor

$$\mu_{\max} = \frac{1}{2\pi N_0} \int_{-\infty}^{\infty} |X(\omega)|^2 d\omega$$



Por último, cabe anotar que la estimación de la señal a la salida del filtro acoplado es sesgada, por cuanto

$$\begin{aligned} E \{ \tilde{x}(t) \} &= E \left\{ \int_0^T h_{ac}(\tau; t_0) z(t - \tau) d\tau \right\} = \int_0^T h_{ac}(\tau; t_0) x(t - \tau) d\tau \\ &= x_{1c}(t) \neq x(t) \end{aligned}$$

**Filtros activos y pasivos.** Sea el dispositivo, que realiza la estimación óptima por el criterio del máximo valor de relación señal/ruido de la forma

$$\tilde{x}(t) = \int_T h_{ac}(\tau) z(t - \tau) d\tau \quad (6.65)$$

siendo  $h_{ac}(\tau)$  la solución no homogénea de la ecuación integral (6.65) y  $z(t)$  la combinación aditiva de la señal útil y ruido, observadas en el intervalo  $(t - T, t)$ . La expresión (6.65) se puede interpretar de dos maneras. En el primer caso, el filtro acoplado es un sistema lineal con parámetros constantes, cuya respuesta a impulso  $h_{ac}(\tau)$  se determina por la forma de la señal y de la función de correlación del ruido, acordes con la descripción (6.61). En el caso particular del ruido blanco Gaussiano, la respuesta a impulso se obtiene de la representación espejo de la señal útil.

Sea un generador, que trabaja por el principio descrito por la solución (6.61), entonces, la estimación  $\tilde{x}(t)$  se obtiene mediante el dispositivo de correlación, en el cual la función  $h_{ac}(\tau)$ , obtenida de un generador local, se multiplica la observación de entrada y retenida, mientras el producto obtenido se integra por todo el intervalo de observación. Esta clase de dispositivos para la extracción de señales se denominan *filtros activos*, a diferencia de un sistema lineal con respuesta a impulso  $h_{ac}(\tau)$ , que se denomina *filtro pasivo*.

**Ejemplo 6.4.** Sea la señal pulso cuadrado

$$x(t) = \begin{cases} a, & -\tau_p/2 \leq t \leq \tau_p/2 \\ 0, & |t| > \tau_p/2 \end{cases}$$

El espectro del pulso corresponde al calculado en el ejemplo 1.2,  $X(\omega) = a\tau_p \text{sinc}(\omega\tau_p/2)$ . En concordancia con (6.63) se obtiene que

$$H_{ac}(\omega; t_0) = k(t_0) a\tau_p |\text{sinc}(\omega\tau_p/2)| / S_n(\omega)$$

De la anterior expresión se observa que la respuesta de frecuencia del filtro acoplado se define sobre un rango infinito de frecuencias, mientras su desfase difiere del la señal original. En el sentido estricto, tal función de transferencia es imposible de realizar, por lo que en la práctica el acople del filtro se limita hasta el lóbulo principal, en el cual se considera se concentra el proceso de la mayor parte de la energía de la señal, mientras su desfase se hace tanto como se pueda igual a 0.

### Problemas

**Problema 6.5.** Sea la señal pulso cuadrado

$$x(t) = \begin{cases} a/2 (e^{j\omega_0 t} - e^{-j\omega_0 t}), & -\tau_p/2 \leq t \leq \tau_p/2 \\ 0, & |t| > \tau_p/2 \end{cases}$$

Hallar la función de transferencia y respuesta impulso del respectivo filtro acoplado.

**Problema 6.6.** Sea la señal pulso cuadrado modulado

$$x(t) = \begin{cases} a/2 (e^{j\omega_0 t} - e^{-j\omega_0 t}), & -\tau_p/2 \leq t \leq \tau_p/2 \\ 0, & |t| > \tau_p/2 \end{cases}$$

Hallar la función de transferencia y respuesta impulso del respectivo filtro acoplado.

**Problema 6.7.** Sea la señal de banda ancha

$$x(t) = \begin{cases} a/2 \cos(\omega_0 t + (d_i - 1)\pi/2), & 0 \leq t \leq T \\ 0, & 0 > t > T \end{cases}$$

Asumiendo que la sucesión de valores  $\{d_i = -1, 1 \in T_d : i = 1, \dots, N\}$ ,  $T = NT_d$ , corresponde a una serie de valores aleatorios, hallar la función de transferencia y respuesta impulso del respectivo filtro acoplado.

**Problema 6.8.** Sea el filtro acoplado, que trabaja por el principio (6.65). Hallar la respuesta a impulso del filtro acoplado, asumiendo que la DEP de la señal útil es

$$S_x(\omega) = \frac{S_0}{1 + (\omega T)^2}$$

mientras la perturbación es RBG.

# Capítulo 7

## Análisis de sistemas estocásticos

El principal instrumento de análisis de los sistemas dinámicos complejos corresponde a los métodos de modelado matemático. En el diseño de sistemas dinámicos, aunque sea conocida la estructura de aleatoriedad del proceso a la entrada, es importante determinar las propiedades de estimación del sistema. Por lo tanto, se hace necesario ajustar el análisis con métodos de la matemática estadística.

### 7.1. Modelos de sistemas dinámicos con entradas aleatorias

Como se analiza en el numeral §1.4, la descripción del funcionamiento de un sistema dinámico mediante el análisis en espacio de estados (diagrama de fase) emplea el método de ecuaciones diferenciales,

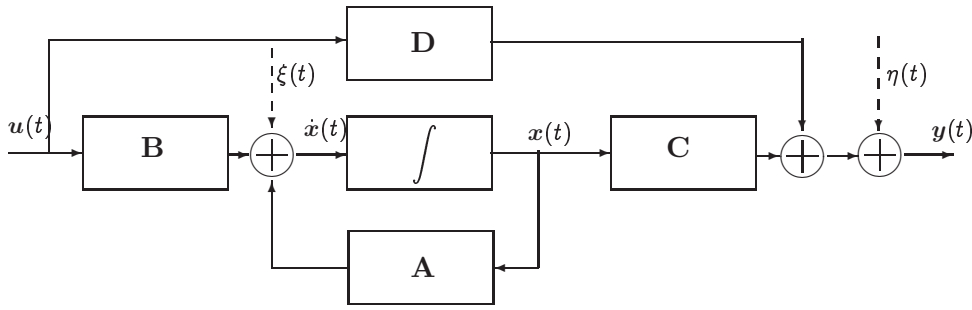
$$\begin{cases} \dot{\mathbf{x}}(t) = \mathbf{f}_x \{ \mathbf{x}(t), \mathbf{u}(t), \boldsymbol{\xi}(t), t \} \\ \mathbf{y}(t) = \mathbf{f}_y \{ \mathbf{x}(t), \mathbf{u}(t), \boldsymbol{\eta}(t), t \} \end{cases} \quad (7.1)$$

donde  $\mathbf{x}_{1 \times n_x}$  es el vector de estados del sistema,  $\mathbf{u}_{1 \times n_u}$  es el vector de la señal de control,  $\mathbf{y}_{1 \times n_y}$  es el respectivo vector de las señales observadas,  $\boldsymbol{\xi}_{1 \times n_\xi}$  es el vector de perturbación en el sistema y  $\boldsymbol{\eta}_{1 \times n_\eta}$  corresponde al vector de errores de medición. Por último,  $t$  es el parámetro escalar de desarrollo del sistema de ecuaciones diferenciales, que típicamente corresponde al tiempo, y el cual toma valores dentro de un intervalo dado,  $t \in T$ .

#### 7.1.1. Sistemas lineales estacionarios

Al asumir la linealidad del sistema en análisis, cuyo diagrama general se muestra en la Figura 7.1, la respectiva descripción (7.1) se simplifica hasta,

$$\begin{cases} \dot{\mathbf{x}}(t) = \mathbf{A}(t) \mathbf{x}(t) + \mathbf{B}(t) \mathbf{u}(t) + \mathbf{I}_\xi(t) \boldsymbol{\xi}(t), \\ \mathbf{y}(t) = \mathbf{C}(t) \mathbf{x}(t) + \mathbf{D}(t) \mathbf{u}(t) + \mathbf{I}_\eta(t) \boldsymbol{\eta}(t), \end{cases} \quad (7.2)$$



**Figura 7.1.** Diagrama general de un sistema dinámico lineal

siendo  $\mathbf{A}(t)$ ,  $\mathbf{B}(t)$ ,  $\mathbf{C}(t)$ ,  $\mathbf{D}(t)$ ,  $\mathbf{I}_\xi(t)$  y  $\mathbf{I}_\eta(t)$  las respectivas matrices de coeficientes de linealidad, que en general, dependen del parámetro  $t$  de desarrollo del sistema. En cuanto a las señales aleatorias,  $\xi(t)$  y  $\eta(t)$ , en los sistemas reales aparecen dificultades relacionadas con la descripción de su influencia en las ecuaciones (7.2).

En el caso simplificado de estudio de un sistema lineal estacionario, la descripción (7.2) se restringe al análisis de las matrices  $\mathbf{A}$ ,  $\mathbf{B}$  y  $\mathbf{C}$  que no dependen del tiempo, luego, para un vector dado de observaciones de salida  $\mathbf{y}$ , el modelo se resume en

$$\begin{cases} \dot{\mathbf{x}} = \mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{u} \\ \mathbf{y} = \mathbf{C}^\top \mathbf{x} \end{cases}$$

con condiciones iniciales  $\mathbf{A}\mathbf{x}(0)$ .

En el numeral §1.4 se analizan tres particularidades de los sistemas lineales, en su descripción mediante el método de ecuaciones diferenciales:

1. Es necesario determinar las condiciones iniciales y las variables de estado,
2. La solución de la ecuación diferencial, que describe el sistema, puede obtenerse mediante un procesador analógico equivalente, el cual se considera el modelo,
3. El modelo del sistema también brinda la posibilidad de realizar el modelado de las señales de salida del sistema.

### 7.1.2. Variables de estado y modelado de procesos aleatorios

La ecuación diferencial, que describe un sistema dinámico lineal, puede ser empleada en el modelado de procesos aleatorios. En el ejemplo 5.1, se muestra el empleo del método de ecuaciones diferenciales en cálculo de los momentos de un circuito  $RC$ . El método de ecuaciones diferenciales ofrece mayores posibilidades para el modelado de procesos aleatorios, como se analiza en el numeral §1.4, la ecuación (1.59) de orden  $m$  puede ser representada en forma alterna por el sistema de ecuaciones de primer orden (1.61) o su variante de descripción vectorial (1.62), a la cual le corresponde la señal de salida (1.64). Del ejemplo 5.1, se puede inferir que este modelo describe procesos, tanto estacionarios como no estacionarios.

Así por ejemplo, en el modelo de la Figura 1.12, sean considerados los correspondientes valores  $t_0 \rightarrow -\infty$ ,  $d^i \eta / dt^i \big|_{t=t_0} = 0$ . Si en calidad de entrada se toma una señal aleatoria vector  $\xi(t)$ , entonces a la salida del sistema se tendrá también otra señal aleatoria  $\eta(t)$ . Ambas componentes vectoriales de las correspondientes señales aleatorias se describen por el sistema (1.70) en la forma,

$$\frac{d\zeta}{dt} = \mathbf{A}(t)\zeta(t) + \mathbf{B}(t)\xi(t), \quad \eta(t) = \mathbf{C}(t)\zeta(t), \quad -\infty < t < \infty$$

En general, si la señal de entrada es ruido blanco Gaussiano, la linealidad del sistema implica un proceso aleatorio de salida también Gaussiano, pero no estacionario. Cabe anotar, que entre más compleja sea la representación de la densidad espectral de potencia que tenga el proceso (mayor grado de aleatoriedad), mayor es la dimensión requerida para el respectivo vector de estado.

**Modelado estocástico.** Se basa en la ecuación diferencial estocástica de la forma,

$$d\mathbf{x} = \mathbf{f}(\mathbf{x}, t) dt + \mathbf{g}(\mathbf{x}, t) d\boldsymbol{\nu} \quad (7.3)$$

donde  $\boldsymbol{\nu}(t)$ ,  $t \in T$ , corresponde a un proceso de Wiener de dimensión  $n$  con función de covarianza de incrementos  $\mathbf{I} dt$ .

Un *proceso escalar de Wiener* se puede definir como un proceso aleatorio  $\nu(t)$ ,  $t \geq t_0$ , con las siguientes propiedades:

1.  $\nu(t_0) = 0$ ,
2.  $\nu(t) \forall t \geq t_0$ , tiene distribución Gaussiana con valor medio igual a cero,
3. El proceso  $\nu(t)$  presenta incrementos estacionarios e independientes, en el sentido en que para cualquier valor de  $t_i \geq t_0$  ( $i = 1, \dots, k$ ), tales que  $t_1 < t_2 < \dots < t_k$ , los correspondientes valores aleatorios:  $\nu(t_1)$ ,  $\nu(t_2) - \nu(t_1)$ ,  $\dots$ ,  $\nu(t_k) - \nu(t_{k-1})$  son mutuamente independientes, mientras la distribución de los incrementos descritos por  $\nu(t) - \nu(s)$ , depende estrictamente de la diferencia de argumentos,  $t - s$ .

De forma similar, se define el proceso de Wiener para el caso de dimensión múltiple, que como consecuencia de las anteriores propiedades, tiene una varianza que crece linealmente en el tiempo, en la forma  $\sigma_\nu^2 = kt$ ,  $k = \text{const.}$ , mientras su función de covarianza se determina mediante la expresión:  $K_\nu(t, s) = k \min(t, s)$ . De (7.3), se tiene que  $\mathbf{x} = \mathbf{x}(t)$  es la solución de la ecuación integral estocástica:

$$\mathbf{x}(t) = \mathbf{x}(t_0) + \int_{t_0}^t \mathbf{f}(\mathbf{x}(\tau), \tau) d\tau + \sum_{t_0}^t \mathbf{g}(\mathbf{x}(\tau), \tau) d\boldsymbol{\nu}(\tau) \quad (7.4)$$

Básicamente, la dificultad en el cálculo de la solución de (7.3) está en la correcta determinación del último término de (7.4). En este sentido, se emplea la integral de Ito, (5.36),

expresada en la forma [11]:

$$\int_{t_0}^t g(\mathbf{x}(\tau), \tau) d\boldsymbol{\nu}(\tau) = \lim_{\Delta t_i \rightarrow 0} \sum_i^N g(\mathbf{x}(t_i), t_i) (\boldsymbol{\nu}(t_{i+1}) - \boldsymbol{\nu}(t_i))$$

siendo  $\Delta t_i = t_{i+1} - t_i$ . La media y covarianza de la solución, respectivamente, son:

$$E\{\mathbf{x}(t)\} = E\{\mathbf{x}(t_0)\} + E\left\{\int_{t_0}^t f(\mathbf{x}(s), s) ds\right\}$$

$$\text{cov}\{\mathbf{x}(t + \varepsilon) - \mathbf{x}(t) | \mathbf{x}(t)\} = g(\mathbf{x}, t) g^T(\mathbf{x}, t) \varepsilon + o(\varepsilon)$$

donde  $o(\varepsilon)$  es un valor de orden infinitamente pequeño con relación a  $\varepsilon$ , tal que  $o(\varepsilon)/\varepsilon \rightarrow 0$  cuando  $\varepsilon \rightarrow 0$ .

Sea la función  $\mathbf{y}(\mathbf{x}, t)$ , que se supone continuamente diferenciable por  $t$  y dos veces continuamente diferenciable por  $\mathbf{x}$ , donde  $\mathbf{x}$  cumple la ecuación (7.3). Entonces,  $\mathbf{y}(\mathbf{x}, t)$  cumple la siguiente ecuación diferencial:

$$\begin{aligned} d\mathbf{y} &= \frac{\partial \mathbf{y}}{\partial t} dt + \sum_{i=1}^n \frac{\partial \mathbf{y}}{\partial x_i} + \frac{1}{2} \sum_{i,j,k=1}^n \frac{\partial^2 \mathbf{y}}{\partial x_i \partial x_j} g_{ik} g_{jk} dt \\ &= \left( \frac{\partial \mathbf{y}}{\partial t} + \sum_{i \neq 1}^n \frac{\partial \mathbf{y}}{\partial x_i} f_i + \frac{1}{2} \sum_{i,j,k=1}^n \frac{\partial^2 \mathbf{y}}{\partial x_i \partial x_j} g_{ik} g_{jk} \right) dt + \sum_{i=1}^n \frac{\partial \mathbf{y}}{\partial x_i} (gd\boldsymbol{\nu})_i \end{aligned} \quad (7.5)$$

siendo  $g_{ij}$  los elementos de la matriz  $\mathbf{g}(\mathbf{x}, t)$  y  $(gd\boldsymbol{\nu})_i$  es el elemento  $i$  del vector  $(gd\boldsymbol{\nu})$ . La ecuación (7.5) presenta ya una forma implementable para la solución de la ecuación diferencial estocástica. Así por ejemplo, de manera particular se analiza para (7.3) la siguiente ecuación lineal diferencial, dada en forma vectorial,

$$d\mathbf{x} = \mathbf{A}(t) \mathbf{x} dt + d\boldsymbol{\nu} \quad (7.6)$$

donde  $\mathbf{x} = \mathbf{x}(t)$  es un vector con dimensión  $n$ ,  $\mathbf{A}(t)$  es una matriz cuadrada de orden  $n \times n$ , mientras  $\boldsymbol{\nu}(t), t \in T$  es un proceso de Wiener de dimensión  $n$  y función de covarianza  $K_{\boldsymbol{\nu}} dt$ . Además, al suponer que el valor inicial  $\mathbf{x}(t_0)$  es un valor aleatorio Gaussiano con valor medio  $\mathbf{m}_{1\mathbf{x}_0}$  y función de covarianza  $K_{\mathbf{x}_0}$ , entonces, la ecuación diferencial estocástica (7.6) se escribe en la forma integral equivalente como

$$\mathbf{x}(t) = \boldsymbol{\Phi}(t, t_0) \mathbf{x}(t_0) + \int_{t_0}^t \boldsymbol{\Phi}(t, s) d\boldsymbol{\nu}(s)$$

La matriz  $\boldsymbol{\Phi}$  cumple la ecuación diferencial (con condiciones iniciales  $\boldsymbol{\Phi}(t_0, t_0) = \mathbf{I}$ ):

$$\frac{d}{dt} \boldsymbol{\Phi}(t, t_0) = \mathbf{A}(t) \boldsymbol{\Phi}(t, t_0)$$

La solución de la ecuación estocástica diferencial (7.6) corresponde a un proceso aleatorio con valor medio  $E\{\mathbf{x}(t)\} = \mathbf{m}_{1\mathbf{x}}(t)$  y función de covarianza  $K_{\mathbf{x}}(s, t)$ , donde

$$\begin{aligned} dm_{1\mathbf{x}}/dt &= \mathbf{A}(t) m_{1\mathbf{x}}(t_0) \quad m_{1\mathbf{x}}(t_0) = m_{1\mathbf{x}_0} \\ K_{\mathbf{x}}(s, t) &= \begin{cases} \Phi(s, t) \mathbf{P}(t), & s \geq t \\ \mathbf{P}(s) \Phi(t, s), & s < t \end{cases} \end{aligned}$$

En cuanto a la matriz  $\mathbf{P}$ , ésta cumple la siguiente ecuación diferencial,

$$\frac{d\mathbf{P}}{dt} = \mathbf{A}\mathbf{P} + \mathbf{P}\mathbf{A}^\top + \mathbf{K}_\nu, \quad \mathbf{P}(t_0) = \mathbf{K}_{\mathbf{x}_0}$$

De otra parte, cuando el conjunto de valores sobre los cuales se determina el intervalo de  $t$  corresponde a los enteros, esto es, se tiene la malla de valores  $T \in \{\dots, -1, 0, 1, \dots\}$ , el sistema determinístico, ya discreto en el tiempo, se describe por la ecuación de iteraciones,

$$\begin{cases} \mathbf{x}[t+1] &= \mathbf{f}_x\{\mathbf{x}[t], \mathbf{u}[t], t\} \\ \mathbf{y}[t] &= \mathbf{f}_y\{\mathbf{x}[t], \mathbf{u}[t], t\} \end{cases}$$

donde  $\mathbf{x}[t]$  es el vector de estados de dimensión  $n$ ,  $\mathbf{y}[t]$  es el vector de observaciones de dimensión  $l$ , mientras  $\mathbf{u}[t]$  es el vector de control. Al agregar factores aditivos de perturbación, el modelo anterior se completa hasta la forma,

$$\begin{cases} \mathbf{x}[t+1] &= \mathbf{f}_x\{\mathbf{x}[t], \mathbf{u}[t], t\} + \mathbf{f}_\xi\{\mathbf{x}[t], \xi[t], t\} \\ \mathbf{y}[t] &= \mathbf{f}_y\{\mathbf{x}[t], \mathbf{u}[t], t\} + \mathbf{f}_\eta\{\mathbf{x}[t], \eta[t], t\} \end{cases}$$

donde  $\mathbf{f}_\xi\{\cdot\}$  y  $\mathbf{f}_\eta\{\cdot\}$  son los funcionales para las perturbaciones del modelo. La simplificación en la solución del modelo, se obtiene en forma práctica, asumiendo la dependencia lineal entre la función, en general, vectorial  $\mathbf{f}_x$  y la señal  $\mathbf{x}[t]$ , además asumiendo que el control del sistema se realiza al escoger adecuadamente los elementos de  $\mathbf{f}_x$ , por lo que la siguiente ecuación estocástica diferencial lineal describe el vector de estados del sistema:

$$\mathbf{x}[t+1] = \Phi[t+1, t] \mathbf{x}[t] + \mathbf{e}[t] \quad (7.7)$$

Sean los vectores  $\mathbf{e}[t]$  y  $\mathbf{e}[s]$  Gaussianos e independientes,  $\forall t \neq s$ , con los momentos:

$$E\{\mathbf{e}[t]\} = 0, \quad E\{\mathbf{e}[t] \mathbf{e}^\top[s]\} = \mathbf{R}_e(t, s)$$

Así mismo, se considera el estado inicial del sistema  $\mathbf{x}[t_0]$  una variable aleatoria Gaussiana con valor medio  $\mathbf{m}_{1\mathbf{x}_0}$  y función de covarianza  $\mathbf{K}_{\mathbf{x}_0}$ . Entonces, la solución de la ecuación lineal estocástica (Gaussiana) de iteraciones (7.7) corresponde a un proceso también Gaussiano con valor medio que cumple la relación (siendo  $\Phi$  alguna matriz de orden  $n \times n$ ):

$$\mathbf{m}_{1\mathbf{x}}[t+1] = \Phi[t+1, t] \mathbf{m}_{1\mathbf{x}}[t]$$

con condición inicial  $\mathbf{m}_{1\mathbf{x}}[t_0]$  y función de covarianza

$$K_{\mathbf{x}}(s, t) = \Phi[s, t] \mathbf{P}[t], \quad \forall s \leq t$$

donde la matriz  $\mathbf{P}(t)$  se halla de la relación iterativa

$$\mathbf{P}[t+1] = \Phi[t+1, t] \mathbf{P}[t] \Phi^T[t+1, t] + \mathbf{K}_{\mathbf{x}}$$

con condiciones iniciales  $\mathbf{P}[t] = \mathbf{K}_{\mathbf{x}_0}$ . Si las matrices  $\Phi$  y  $\mathbf{K}_{\mathbf{x}}$  son constantes en el tiempo, entonces, a partir de todas las anteriores condiciones derivadas del modelo (7.7), se tiene,

$$\mathbf{P}[t] = \Phi^T \mathbf{K}_{\mathbf{x}_0} (\Phi^T)^t + \sum_{s=0}^{t-1} \Phi^s \mathbf{K}_{\mathbf{x}} (\Phi^T)^s$$

**Modelado discreto de sistemas estocásticos continuos.** En este caso, se tiene el cambio de las ecuaciones estocásticas diferenciales a las respectivas ecuaciones de iteración.

Sean los procesos, de funcionamiento del sistema  $\mathbf{x}(t)$  y de obtención de las observaciones  $\mathbf{y}(t)$ , descritos por las ecuaciones estocásticas diferenciales vectoriales de la forma,

$$\begin{cases} d\mathbf{x} = \mathbf{A}\mathbf{x}dt + d\mathbf{v}_1 \\ d\mathbf{y} = \mathbf{C}\mathbf{x}dt + d\mathbf{v}_2 \end{cases} \quad (7.8)$$

donde  $\mathbf{x}_{1 \times n}$  es el vector de estados del sistema, mientras  $\mathbf{y}_{1 \times r}$  es el vector de todas las observaciones,  $\{\mathbf{v}_m(t) : m = 1, 2\}$ , son los vectores los procesos de Wiener con dimensiones respectivas  $n$  y  $r$ , y funciones de covarianza de incremento:  $\mathbf{K}_{\mathbf{v}_m}dt$ . Se asume, además, que las variables de salida se observan a lo largo de la malla de valores discretos de tiempo  $\{t_k : k = 0, 1, \dots\}$ . Entonces, los valores de las variables de estado y de las variables de salida observadas para las ecuaciones diferenciales estocásticas (7.8), sobre una malla discreta de tiempo  $\{t_k : k = 0, 1, \dots\}$ , están relacionadas por las ecuaciones de diferencias estocásticas:

$$\begin{cases} \mathbf{x}(t_{i+1}) = \Phi \mathbf{x}(t_i) + \hat{\mathbf{v}}_1(t_i) \\ \mathbf{z}(t_{i+1}) = \mathbf{y}(t_{i+1}) - \mathbf{y}(t_i) = \mathbf{S}\mathbf{x}(t_i) + \hat{\mathbf{v}}_2(t_i) \end{cases}$$

donde la matriz  $\Phi = \Phi(t_{i+1}, t_i), \forall i \geq 0$ , se determina por las relaciones

$$\frac{d}{dt} \Phi(t, t_i) = \mathbf{A}(t) \Phi(t, t_i), \quad \Phi(t_i, t_i) = \mathbf{I}, \quad t_i \leq t \leq t_{i+1},$$

mientras, la matriz  $\mathbf{S} = \mathbf{S}(t_{i+1}, t_i), \forall i \geq 0$ , se da por la expresión,

$$\mathbf{S}(t_{i+1}, t_i) = \int_{t_i}^{t_{i+1}} \mathbf{C}(s) \Phi(s, t_i) ds \quad (7.9)$$

Los vectores  $\{\hat{\mathbf{v}}_m(t_i) : m = 1, 2; \forall i \geq 0\}$  corresponden a sucesiones de valores aleatorios



Gaussianos independientes con valor medio igual a cero y con las correspondientes funciones de correlación, determinadas a partir de las siguientes expresiones ( $\forall i \geq 0$ ):

$$\mathbf{R}_{\boldsymbol{\nu}_1}(t_i) = \int_{t_i}^{t_{i+1}} \boldsymbol{\Phi}(t_{i+1}, s) \mathbf{R}_{\boldsymbol{\nu}_1} \boldsymbol{\Phi}^T(t_{i+1}, s) ds \quad (7.10a)$$

$$\mathbf{R}_{\boldsymbol{\nu}_2}(t_i) = \int_{t_i}^{t_{i+1}} \left( \mathbf{S}(t_{i+1}, s) \mathbf{R}_{\boldsymbol{\nu}_1} \mathbf{S}^T(t_{i+1}, s) + \mathbf{R}_{\boldsymbol{\nu}_2}(s) \right) ds \quad (7.10b)$$

Finalmente, la función de covarianza mutua entre los procesos  $\{\boldsymbol{\nu}_m : m = 1, 2\}$  se determina por la expresión,

$$\mathbf{K}_{\boldsymbol{\nu}_1 \boldsymbol{\nu}_2}(t_i) = E \left\{ \boldsymbol{\nu}_1(t_i) \boldsymbol{\nu}_2^T(t_i) \right\} = \int_{t_i}^{t_{i+1}} \boldsymbol{\Phi}(t_i, s) \mathbf{R}_{\boldsymbol{\nu}_1}(s) \mathbf{S}^T(t_{i+1}, s) ds, \quad i \geq 0$$

Las ecuaciones iterativas estocásticas (7.9), (7.10a), (7.10b), mas las ecuaciones diferenciales estocásticas (7.8), desde el punto de vista de sus propiedades estadísticas dentro del intervalo de discretización, son idénticas, con lo cual el respectivo proceso de funcionamiento continuo se puede simular en los sistemas de proceso digital.

### 7.1.3. Estimación de parámetros en sistemas estocásticos

La necesidad de aumentar la efectividad del funcionamiento de los sistemas de control, en condiciones de factores aleatorios, estimuló el desarrollo de métodos de optimización orientada a mejorar el registro y precisión de la información sobre las propiedades de la planta a diseñar, así como de los regímenes de su funcionamiento.

El trabajo de un sistema se debe representar por algún modelo, por ejemplo, los descritos en el numeral §7.1, en el cual, el vector de estados registra las medidas de los diferentes dispositivos de captura de información sobre el estado del sistema, dicho de otra manera, es el vector de mediciones.

Sea el objeto de análisis un sistema dinámico en tiempo discreto, que se describe por las siguientes ecuaciones de diferencia:

$$\mathbf{x}[k+1] = \mathbf{A}\mathbf{x}[k], \quad k \geq 0, \quad k \in \mathbb{Z} \quad (7.11)$$

siendo  $\mathbf{x}$  un vector con dimensión  $n$  y  $\mathbf{A}$  una matriz cuadrada de orden  $n \times n$ . Se asume que el proceso de medición se describe por

$$\mathbf{y}[k] = \mathbf{C}\mathbf{x}[k] \quad (7.12)$$

donde  $\mathbf{y}$  es el vector de medidas con dimensión  $r$ , mientras la matriz  $\mathbf{C}$  tiene orden  $r \times n$ .

La planta se considera *observable* si a partir de los registros de medida  $\mathbf{y}[0], \dots, \mathbf{y}[n-1]$  se puede determinar el estado  $\mathbf{x}[0]$  y, por lo tanto, todos los demás subsecuentes estados

del sistema,  $\mathbf{x}[1], \mathbf{x}[2], \dots$

Al describir de forma secuencial la ecuación (7.12), para los valores de  $k = 0, \dots, n-1$  se obtiene la siguiente expresión matricial,

$$\begin{bmatrix} \mathbf{y}^T[0] : \mathbf{y}^T[1] : \dots : \mathbf{y}^T[n-1] \end{bmatrix} = \mathbf{x}^T[0] \mathbf{H} \quad (7.13)$$

donde  $\mathbf{H} = [\mathbf{C}^T : \mathbf{A}^T \mathbf{C}^T : \dots : \mathbf{A}^{T[n-1]} \mathbf{C}^T]$ .

En general, la condición de observabilidad se puede hacer equivalente a la condición de existencia y unicidad de la solución  $\mathbf{x}[0]$  en (7.13), que consiste en que el rango de la matriz  $\mathbf{H}$  es igual a  $n$ . Entonces, la pareja de matrices  $\mathbf{A}, \mathbf{C}$  se denomina *observable*.

El concepto de *identificación*, que está asociado con la determinación de la matriz  $\mathbf{A}$  (dados los valores medidos del vector de estados), es equivalente a la condición de existencia y unicidad de la solución  $\mathbf{A}$  de la ecuación matricial,

$$\begin{bmatrix} \mathbf{x}[1] : \dots : \mathbf{x}[n] \end{bmatrix} = \mathbf{A} \begin{bmatrix} \mathbf{x}[0] : \dots : \mathbf{A}^{n-1} \mathbf{x}[0] \end{bmatrix} = \mathbf{A} \mathbf{S}$$

lo que implica que la matriz  $\mathbf{S}$  tenga rango  $n$ .

Las anteriores condiciones pueden ser generalizadas para el caso de un sistema dinámico en tiempo continuo, para todo valor  $t_0 \leq t \leq T$ , cuyo funcionamiento y proceso de medida se dan por las respectivas ecuaciones diferenciales,

$$\begin{cases} \dot{\mathbf{x}}(t) = \mathbf{A}(t) \mathbf{x}(t) \\ \mathbf{y}(t) = \mathbf{C}(t) \mathbf{x}(t) \end{cases}$$

Si se nota la matriz de transición del sistema lineal  $\dot{\mathbf{x}}(t) = \mathbf{A}(t) \mathbf{x}(t)$  por  $\Phi(\tau, t)$ , entonces la condición de observabilidad consiste en el cumplimiento de la siguiente relación:

$$\mathbf{M}(t, t_0) = \int_{t_0}^t \Phi^T(\tau, t) \mathbf{C}^T(\tau) \mathbf{C}(\tau) \Phi(\tau, t) d\tau > 0$$

La *matriz de observabilidad*  $\mathbf{M}(t, t_0)$ , mediante la cual se puede determinar el vector  $\dot{\mathbf{x}}(t)$ :  $\mathbf{x}(t) = \mathbf{M}^{-1}(t, t_0) \mathbf{y}(t)$ , se halla de la ecuación diferencial,

$$\frac{d}{dt} \mathbf{M}(t, t_0) = -\mathbf{A}^T(t) \mathbf{M}(t, t_0) - \mathbf{M}(t, t_0) \mathbf{A}(t) + \mathbf{C}^T(t) \mathbf{C}(t)$$

con condición inicial  $\mathbf{M}(t, t_0) = \mathbf{0}$ .

De forma similar, la anterior descripción de la condición de observabilidad se mantiene para los sistemas dinámicos en tiempo discreto. Sea  $\Phi[i, j]$  la matriz de transición del sistema (7.11), tal que,

$$\mathbf{x}[k] = \Phi[k, 0] \mathbf{x}[0], \quad \Phi[0, 0] = \mathbf{I}$$

Si se formula la matriz de observabilidad en la forma,

$$\mathbf{M}[k, 0] = \sum_{l=0}^k \Phi^T[l, k] \mathbf{C}^T[l] \mathbf{C}[l] \Phi[l, k]$$

entonces, la condición de observabilidad se escribe en forma de la desigualdad:  $\mathbf{M}[k, 0] > 0$ .

En la práctica, durante el proceso de registro de señales ocurren errores aleatorios de medida, así mismo, el funcionamiento del propio sistema dinámico se ve afectado por perturbaciones, también de naturaleza aleatoria. Además, la cantidad de registros de medida a procesar pueden ser insuficientes, o por el contrario, estar en número excesivo. Por lo tanto, el principal problema consiste en la obtención de las estimaciones óptimas para las variables que describen, tanto los parámetros, como el estado del sistema dinámico. Asociada a la estimación óptima de los valores de medida está la filtración, que se analiza en el capítulo §6. Sin embargo, también relacionada con la tarea de filtración está el problema de la determinación de los parámetros del sistema o *identificación*. Aunque son suficientemente conocidas las diversas técnicas de filtración óptima de sistemas dinámicos, su efectividad está condicionada por la correspondencia que existe entre los parámetros del modelo y la planta real de análisis, que no siempre es suficientemente precisa. En este sentido, se han mejorado las técnicas clásicas de filtración para el caso de datos a priori dados de forma no adecuada, que nutren los algoritmos de filtración. El mejoramiento de los algoritmos se puede alcanzar con el empleo de técnicas estables de estimación, basadas en algoritmos que se adaptan a las condiciones de cambio externas.

#### 7.1.4. Identificación de parámetros en sistemas lineales

Uno de los métodos más efectivos, al mismo tiempo más cómodos de representación en sistemas de proceso digital, es el *método de Kaczmarz* para la identificación de parámetros en sistemas de dimensión múltiple. En particular, sea el sistema lineal discreto en la forma,

$$\mathbf{y}[k] = \mathbf{c}^T \mathbf{x}[k], \quad k = 1, 2, \dots \quad (7.14)$$

donde  $\mathbf{y}[k]$  es el escalar de la señal de salida del sistema,  $\mathbf{x}[k]$  es el vector de la señal de entrada y  $\mathbf{c}$  es un vector de parámetros desconocidos del sistema ( $\mathbf{c}$  y  $\mathbf{x}$  con dimensión  $m$ ), de tal manera, que se pueden estimar los valores del vector de parámetros  $\mathbf{c}$  del sistema a partir de las sucesiones de observaciones  $\{\mathbf{x}[k] : k > 1\}$  y  $\{\mathbf{y}[k] : k > 1\}$ . La efectividad del método principalmente depende de las cualidades estocásticas de la sucesión de entrada  $\{\mathbf{x}[k]\}$ . Así por ejemplo, bajo ciertas presunciones y asumiendo la ortogonalidad mutua de los vectores  $\mathbf{x}[1], \dots, \mathbf{x}[m]$ , entonces el valor del vector  $\mathbf{c}$  puede ser obtenido con suficiente precisión en solo  $m$  pasos de cálculo. Además, si los mismos vectores, siendo procesos aleatorios, son Gaussianos con igual varianza, entonces la estimación del vector  $\mathbf{c}$  demanda una muestra de volumen  $l = (4 \dots 5)$  términos.

El método de Kaczmarz tiene amplio uso en la solución de problemas de identificación de sistemas lineales no estacionarios y su algoritmo de cálculo tiene estructura simple, basada

en las siguientes relaciones recurrentes:

$$\tilde{\mathbf{c}}[k] = \tilde{\mathbf{c}}[k-1] + \frac{\mathbf{y}[k] - \tilde{\mathbf{c}}^T[k-1] \mathbf{x}[k]}{\mathbf{x}^T[k] \mathbf{x}[k]} \mathbf{x}[k], \quad k = 2, 3, \dots \quad (7.15)$$

donde  $\tilde{\mathbf{c}}[k]$  es la estimación del vector de parámetros desconocidos en el paso  $k$ . El valor inicial del vector  $\tilde{\mathbf{c}}[1]$  se fija de manera a priori.

La interpretación geométrica de la relación (7.15) consiste en que cada estimación  $\tilde{\mathbf{c}}[k]$  puede considerarse como la proyección del valor previamente calculado de  $\tilde{\mathbf{c}}[k-1]$  sobre el respectivo  $k$  hiperplano en un espacio euclídeo  $\mathbb{R}^m$ , determinado por la relación (7.14). En este caso, la sucesión de las normas de los vectores  $\Delta \mathbf{c}[k] = \tilde{\mathbf{c}}[k] - \mathbf{c}[k]$  converge, decreciendo de forma monótona, hasta cero. Sin embargo, la presunción de independencia estadística, implica que en la práctica, al aumentar la correlación entre valores cercanos en el tiempo de la sucesión de entrada  $\mathbf{x}[1], \dots, \mathbf{x}[m]$ , de manera significativa disminuye la efectividad del algoritmo de Kaczmarz. En este sentido, se han propuesto varias modificaciones al algoritmo básico (7.15), que además permiten aumentar su velocidad de proceso [45, 46].

El algoritmo generalizado de estimación recurrente de Kaczmarz se puede describir por las siguientes expresiones recurrentes:

$$\begin{aligned} \tilde{\mathbf{c}}[k] &= \tilde{\mathbf{c}}[k-1] + \frac{\mathbf{y}[k] - \tilde{\mathbf{c}}^T[k-1] \mathbf{x}[k]}{\mathbf{x}^T[k] \mathbf{x}[k]} \mathbf{x}[k], \quad k = 2, 3, \dots \\ \hat{\mathbf{c}}[k] &= \tilde{\mathbf{c}}[k] + \alpha[k] (\hat{\mathbf{c}}[k-2] - \hat{\mathbf{c}}[k-1]) \end{aligned}$$

siendo  $\hat{\mathbf{c}}[k]$  la estimación actual del vector de parámetros  $\mathbf{c}$  y  $\alpha$  el factor de corrección, que se calcula como,

$$\alpha[k] = \begin{cases} 1, & \rho^2 \{\hat{\mathbf{c}}[k-2], \hat{\mathbf{c}}[k]\} > (\rho^2 \{\hat{\mathbf{c}}[k-2], \hat{\mathbf{c}}[k-1]\} - \mu)^2 \\ 0, & \text{caso contrario} \end{cases}$$

donde

$$\rho\{\hat{\mathbf{c}}[m], \hat{\mathbf{c}}[n]\} = \frac{\mathbf{y}[n] - \mathbf{c}^T[m] \mathbf{x}[n]}{\mathbf{x}^T[n] \mathbf{x}[n]}$$

El factor  $\mu$  es igual a

$$\mu = \mathbf{y}[k] - \hat{\mathbf{c}}^T[k-1] \mathbf{x}[k] / \mathbf{x}^T[k] \mathbf{x}[k]$$

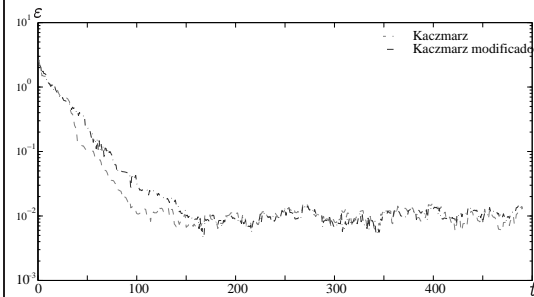
para el caso del algoritmo [46] y  $\mu = 0$  para el algoritmo presentado en [45].

De esta manera, el algoritmo de Kaczmarz consiste en el procedimiento recurrente con los valores iniciales desconocidos  $\hat{\mathbf{c}}[1]$ ,  $\hat{\mathbf{c}}[2]$  y  $\hat{\mathbf{c}}[2]$ , los cuales deben ser fijados a priori.

**Ejercicio en el CP 7.1.** La identificación de los coeficientes de un filtro no recursivo de orden  $N$ , se puede realizar mediante el algoritmo de Kaczmarz (en sus versiones simple o modificada), generando los respectivos vectores de señal entrada y salida del filtro, como se muestra a continuación:

```
function c = Kaczmarz(x,y,N)
% Algoritmo de Kaczmarz
% Entrada x, salida y, orden N
c = ones(N);
for i=N:length(x),
    ye = c'*x(i-N+1:i);
    R = x(i-N+1:i)'*x(i-N+1:i);
    c = c + ( y(i) - ye ) * R^-1 * x(i-N+1:i);
end;

function c = KaczmarzMod(x,y,N)
% Algoritmo de Kaczmarz modificado
c = ones(N); c1 = ones(N); c2 = ones(N);
for i=N+1:L,
    c2 = c1;    c1 = c;
    R = x(i-N+1:i)'*x(i-N+1:i);
    mu = (y(i) - we_1'*x(i-N+1:i))*R^-1;
    rho_1 = (y(i) - we_2'*x(i-N+1:i))*R^-1;
    rho_2 = (y(i-1) - we_2'*x(i-N:i-1))
    x0 = (x(i-N:i-1)'*x(i-N:i-1))^-1;
    rho_2 = rho_2*x0;
    if rho_1^2 > (rho_2^2-mu)^2,
        alpha = 1;
    else
        alpha = 0;
    end;
    c0 = ( y(i) - c1'*x(i-N+1:i) );
    c = c1 + c0*R^-1*x(i-N+1:i);
    c = c + alpha*( c2 - c1 );
end;
```



La convergencia de ambos procedimientos del algoritmo de Kaczmarz, para el ejemplo de identificación de un filtro de orden  $N = 10$ , se muestra en la Figura de la parte derecha.

## Problemas

**Problema 7.1.** Mediante el análisis de los polos de las respectivas funciones de transferencia  $H(z)$ , comprobar para qué valores de  $k \in [0, 1]$  se cumple la condición de estabilidad de cada uno de los siguientes filtros:

$$H(z) = \frac{1 - z^{-1}}{1 - kz^{-1}}$$

$$H(z) = \frac{1 - z^{-2}}{1 - kz^{-1}}$$

$$H(z) = \frac{50 - 76.5z^{-1} - 10.575z^{-2} + 26.581z^{-3} - 5.882z^{-4}}{1 + 0.8z^{-1} - 0.9z^{-2} - kz^{-3} + 0.9144z^{-4} + 0.5184z^{-5}}$$

$$H(z) = \frac{1 - kz^{-1}}{1 - z^{-1}}$$

$$H(z) = \frac{1 - z^{-2}}{1 - 1.8z^{-1} + kz^{-2}}$$

**Problema 7.2.** Estimar los coeficientes de la función respuesta a impulso de los filtros pasabajos, descritos en los problemas 1.17 y 1.18, empleando los algoritmos de Kaczmarz (simple y modificado). Hallar el error absoluto de aproximación.



## 7.2. Filtración de sistemas dinámicos

### 7.2.1. Filtración Wiener

Sea la entrada de un sistema, el proceso vectorial aleatorio,  $\mathbf{x}(t)$ , de dimensión  $n$  y con valor medio 0, el cual se debe restablecer de la manera más precisa posible, a partir de la estimación  $\hat{\mathbf{x}}(t)$ , que se asume puede ser descrita por el siguiente sistema de ecuaciones diferenciales con coeficientes variables en el tiempo:

$$\frac{d\hat{\mathbf{x}}}{dt} = \mathbf{P}(t)\hat{\mathbf{x}}(t) + \mathbf{Q}(t)\mathbf{z}(t) \quad (7.16)$$

donde  $\mathbf{z}(t)$  es la entrada de dimensión  $n$  del sistema, (7.16), la cual corresponde a un proceso aleatorio del tipo no estacionario con valor medio igual a 0, el vector  $\hat{\mathbf{x}}(t)$  es de dimensión  $n$ , mientras las matrices  $\mathbf{P}(t)$  y  $\mathbf{Q}(t)$  son de orden  $n \times n$  y  $n \times m$ , respectivamente.

El error de representación se determina por la diferencia simple entre la señal deseada y la observación para su estimación,

$$\mathbf{e}(t) = \mathbf{x}(t) - \hat{\mathbf{x}}(t) \quad (7.17)$$

Sin embargo, la calidad del filtro (7.16) se estima por el criterio,

$$J(t) = E \left\{ \mathbf{e}^T(t) \mathbf{e}(t) \right\} \quad (7.18)$$

que se minimiza para optimizar el desempeño del sistema. La parte derecha de (7.18) es la suma de los elementos diagonales de la respectiva matriz,  $\text{trace}\{E\{\mathbf{e}^T(t) \mathbf{e}(t)\}\}$ .

De otra parte, teniendo en cuenta (7.16), la señal  $\mathbf{x}(t)$  se determina mediante la relación,

$$\hat{\mathbf{x}}(t) = \int_{t_0}^t \boldsymbol{\mu}(t, \tau) \mathbf{z}(\tau) d\tau$$

siendo  $\boldsymbol{\mu}(t, \tau)$  una matriz de orden  $n \times m$ , determinada por el producto,

$$\boldsymbol{\mu}(t, \tau) = \boldsymbol{\Phi}(t, \tau) \mathbf{Q}(\tau) \quad (7.19)$$

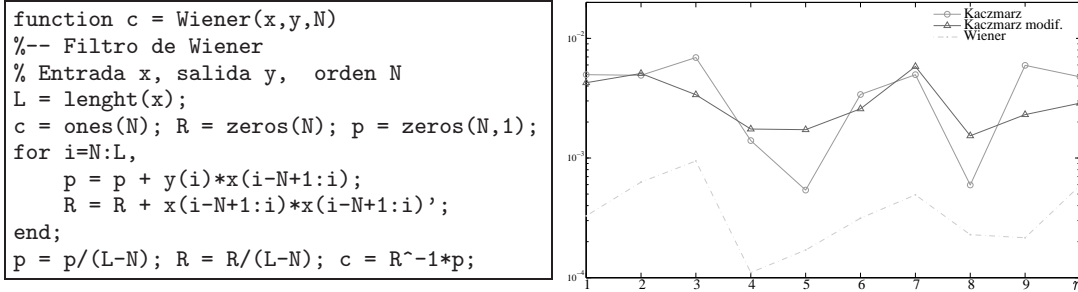
donde  $\boldsymbol{\Phi}(t, \tau) = \mathbf{f}(t) \mathbf{f}^{-1}(\tau)$ ,  $\mathbf{f}$  es la matriz fundamental de soluciones del sistema de ecuaciones diferenciales homogéneas en la forma,  $d\hat{\mathbf{x}}/dt = \mathbf{P}(t)\hat{\mathbf{x}}(t)$ , con condiciones iniciales 0. La selección óptima de la matriz  $\boldsymbol{\mu}(t, \tau)$  de (7.19), que corresponde a la función de peso del filtro (7.16), se minimiza por el criterio  $J(t)$  en (7.18).

Cabe anotar, que la condición de optimalidad de  $J(t)$  es equivalente a la ecuación integral matricial de Wiener-Hopf [47]:

$$E \left\{ \mathbf{x}(t) \mathbf{z}^T(t_1) \right\} = \int_{t_0}^t \boldsymbol{\mu}(t, t_2) E \left\{ \mathbf{z}(t_2) \mathbf{z}^T(t_1) \right\} dt_2, \quad t_0 \leq t_1 \leq t \quad (7.20)$$

En la práctica, la solución de la ecuación integral (7.20), además de exigir el conocimiento a priori de la estructura estocástica de las señales (funciones de correlación y espectros de potencia), demanda grandes recursos computacionales.

**Ejercicio en el CP 7.2.** La identificación del sistema, en el ejemplo 7.1, se realiza mediante el filtro de Wiener, empleando el procedimiento seguidamente ilustrado. Los resultados en el error de aproximación obtenidos en los casos de los algoritmos de Kaczmarz y de Wiener se muestran en la Figura de la parte derecha para valor de  $N = 10$ .



### 7.2.2. Filtración óptima de sistemas discretos lineales

La estimación del vector de estados del sistema sobre el conjunto de observaciones a la salida del sistema, teniendo en cuenta las perturbaciones presentes y los errores de medida, conlleva a una tarea de filtración, que en estos casos, uno de los métodos más efectivos corresponde a la *filtración de Kalman* [48].

Sea el modelo de un sistema dinámico lineal en tiempo discreto,

$$\mathbf{x}[k+1] = \mathbf{A}[k]\mathbf{x}[k] + \mathbf{w}[k], \quad k = 0, 1, \dots \quad (7.21)$$

donde  $\mathbf{x}[k]$  es el vector de estados del sistema que no puede ser medido de forma directa,  $\mathbf{w}[k]$  son las perturbaciones aleatorias con estructura Gaussiana, siendo  $\mathbf{x}$  y  $\mathbf{w}$  de dimensión  $n$ . La matriz determinística  $\mathbf{A}[k]$  es de orden  $n \times n$ .

Las siguientes características del modelo se consideran fijadas a priori:

$$E\{\mathbf{x}[0]\} = \mathbf{x}[0], \quad E\{(\mathbf{x}[0] - \bar{\mathbf{x}}[0])(\mathbf{x}[0] - \bar{\mathbf{x}}[0])^T\} = \mathbf{P}[0]$$

$$E\{\mathbf{w}[k]\} = \mathbf{w}[k], \quad E\{(\mathbf{w}[k] - \bar{\mathbf{w}}[k])(\mathbf{w}[k] - \bar{\mathbf{w}}[k])^T\} = \mathbf{Q}[k]\delta_{kl}$$

$$E\{(\mathbf{w}[k] - \bar{\mathbf{w}}[k])(\mathbf{x}[0] - \bar{\mathbf{x}}[0])^T\} = 0$$

siendo  $\mathbf{Q}[k]$  una matriz conocida.

Además, se considera que la relación, que describe el proceso de medida del vector de



estados, es también lineal,

$$\mathbf{y}[k] = \mathbf{C}[k] \mathbf{x}[k] + \mathbf{v}[k], \quad k = 0, 1, \dots$$

donde  $\mathbf{y}[k]$  es el vector de medida de las respectivas observaciones,  $\mathbf{v}[k]$  es el vector aleatorio Gaussiano de error de medida, ambos vectores  $\mathbf{v}$  e  $\mathbf{y}$  tienen dimensión  $r$  y  $\mathbf{C}[k]$  es una matriz dada de orden  $r \times n$ . Del vector  $\mathbf{v}[k]$ , se considera conocida su función de correlación propia  $\mathbf{R}_v[k]$ , además, se asumen sus siguientes momentos:

$$\begin{aligned} E\{\mathbf{v}[k]\} &= \bar{\mathbf{v}}[k], \quad E\{\mathbf{v}[k] \mathbf{v}^T[k]\} = \mathbf{R}_v[k] \delta_{kl} \\ E\{(\mathbf{w}[k] - \bar{\mathbf{w}}[k]) \mathbf{v}^T[k]\} &= 0, \quad E\{(\mathbf{x}[k] - \bar{\mathbf{x}}[k]) \mathbf{v}^T[k]\} = 0, \end{aligned}$$

Los valores determinados anteriormente, implican que el vector  $\mathbf{v}[k]$  tiene correlación nula, tanto con las perturbaciones del sistema, como con su vector inicial de condiciones. A efectos de simplificar el modelo, sin pérdidas de generalidad, se puede considerar que  $\mathbf{v}[k] = \mathbf{w}[k] = 0$ , mientras las condiciones iniciales se pueden asumir  $\mathbf{x}[0] = 0$ .

Basados en la sucesión de observaciones  $\mathbf{y}[0], \dots, \mathbf{y}[k]$  es necesario hallar la estimación insesgada  $\hat{\mathbf{x}}[k]$ , cuya error  $\mathbf{e}[k]$  se da por la expresión (7.17) cuando el criterio de calidad de la estimación es el valor cuadrático medio,  $E\{\mathbf{e}^2[k]\}$ . En calidad de estimación óptima del vector  $\hat{\mathbf{x}}[k]$  se toma su esperanza matemática condicional [48]:

$$\hat{\mathbf{x}}[k] = E\{\mathbf{x}[k] | \mathbf{y}[0], \dots, \mathbf{y}[k]\}$$

con la cual se obtiene el mínimo valor cuadrático medio del error de estimación.

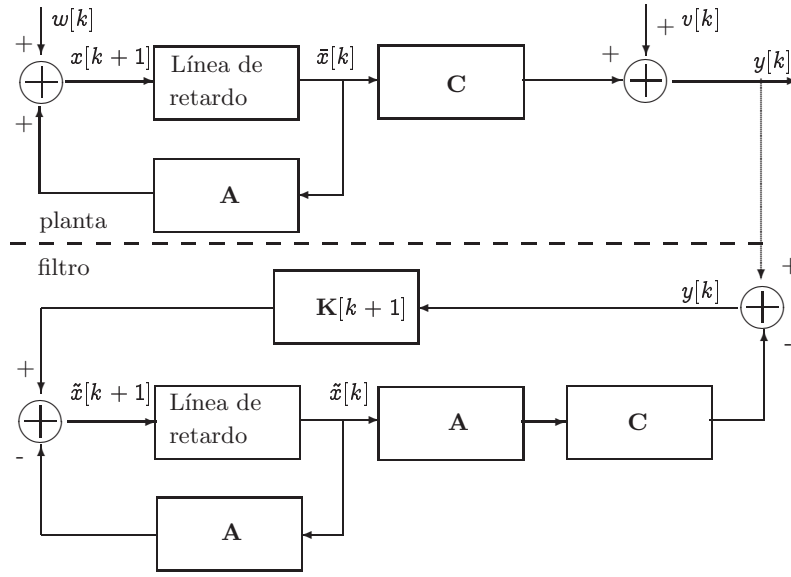
El algoritmo de Kalman, que se describe de forma recurrente, determina el proceso de evolución en el tiempo de la estimación óptima  $\hat{\mathbf{x}}[k]$ :

$$\begin{cases} \hat{\mathbf{x}}[k] &= \mathbf{A}[k-1] \hat{\mathbf{x}}[k-1] \\ \tilde{\mathbf{x}}[k] &= \hat{\mathbf{x}}[k] + \mathbf{K}[k] (\mathbf{y}[k] - \mathbf{C}[k] \hat{\mathbf{x}}[k]), \quad k = 1, 2, \dots \end{cases} \quad (7.22)$$

donde  $\hat{\mathbf{x}}[0]$  se asume conocido. Las demás expresiones recurrentes necesaria en (7.22) se determinan como:

$$\begin{cases} \bar{\mathbf{P}}[k] &= \mathbf{A}[k-1] \hat{\mathbf{P}}[k-1] \mathbf{A}^T[k-1] + \mathbf{Q}[k-1], \\ \mathbf{K}[k] &= \bar{\mathbf{P}}[k] \mathbf{C}^T[k] (\mathbf{C}[k] \bar{\mathbf{P}}[k] \mathbf{C}^T[k] + \mathbf{R}[k])^{-1}, \\ \hat{\mathbf{P}}[k] &= \bar{\mathbf{P}}[k] - \mathbf{K}[k] \mathbf{C}[k] \bar{\mathbf{P}}[k] = (\mathbf{I} - \mathbf{K}[k] \mathbf{C}[k]) \bar{\mathbf{P}}[k], \quad \bar{\mathbf{P}}[0] = \mathbf{P}[0] \end{cases}$$

El diagrama funcional del filtro de Kalman se ilustra en la Figura 7.2, en la cual se muestra que el procedimiento consiste en que la estimación de  $\mathbf{x}[k]$  corresponde a la estimación extrapolada del estado del sistema después de  $k-1$  pasos de medida. La estimación  $\tilde{\mathbf{x}}[k]$  es la corrección del valor estimado  $\hat{\mathbf{x}}[k]$ , obtenido como resultado del proceso de la  $k$  medida. A su vez, las matrices  $\mathbf{P}[k]$  y  $\tilde{\mathbf{P}}[k]$  corresponden, respectivamente a la matriz extrapolada y a la corregida, de las funciones de covarianza de los errores de filtración. Por



**Figura 7.2.** Diagrama funcional del filtro Kalman.

cierto, la matriz de covarianza de los errores de estimación  $\tilde{\mathbf{P}}[k]$  no depende de la sucesión de medidas  $\mathbf{y}[0], \dots, \mathbf{y}[k]$ , por lo tanto, sus valores pueden ser calculados de antemano y llevados a alguna posición de memoria.

El modelo del sistema dinámico de control se puede hacer un tanto más complejo, en la siguiente forma,

$$\mathbf{x}[k+1] = \mathbf{A}[k]\mathbf{x}[k] + \mathbf{B}[k]\mathbf{u}[k] + \mathbf{w}[k]$$

En este caso, la sucesión  $\mathbf{y}[0], \dots, \mathbf{y}[k]$  se determina a partir de la expresión

$$\mathbf{y}[k] = \mathbf{C}[k]\mathbf{x}[k] + \mathbf{v}[k]$$

En cuanto a la sucesión  $\mathbf{u}[0], \dots, \mathbf{u}[k]$ , ésta corresponde a la sucesión determinística dada de control.

De esta manera, el filtro de Kalman se determina por las siguientes relaciones recurrentes:

$$\tilde{\mathbf{x}}[k] = \mathbf{A}[k-1]\tilde{\mathbf{x}}[k-1] + \mathbf{B}[k-1]\mathbf{u}[k-1]$$

$$\hat{\mathbf{x}}[k] = \tilde{\mathbf{x}}[k] + \mathbf{K}[k](\mathbf{y}[k] - \mathbf{C}[k]\tilde{\mathbf{x}}[k])$$

El vector  $\mathbf{x}[0]$  es conocido de antemano y la matriz  $\mathbf{K}$  se calcula, tal cual se hace para el procedimiento descrito anteriormente.

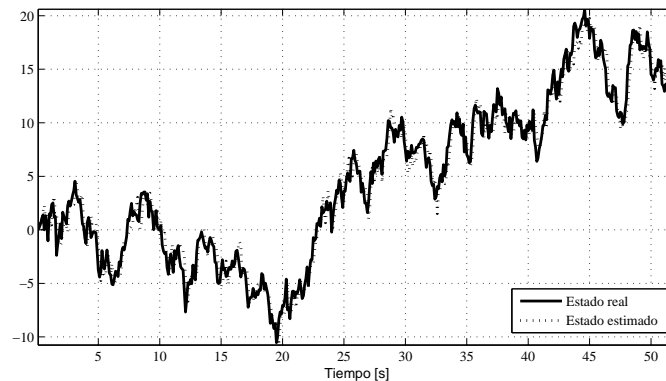
**Ejercicio en el CP 7.3.** Desarrollar el algoritmo de estimación mediante el filtro Kalman para un proceso que describe una caminata aleatoria en la forma:

$$\begin{aligned}x[k+1] &= x[k] + w[k], \\ y[k] &= x[k] + v[k]\end{aligned}$$

siendo las señales aleatorias  $w$  y  $v$ , RBG con momentos:

$$\begin{aligned}E\{w\} &= E\{v\} = 0 \\ \sigma_v &= \sigma_w = 1\end{aligned}$$

```
% filtro de Kalman en un sistema discreto
%--- Inicialización de parámetros de simulación
A = 1;           % Matriz de transición de estado A
C = 1;           % Matriz de medición de estado C
R = 1;           % Matriz de covarianza del ruido de medición
Q = 1;           % Matriz de covarianza de la perturbación del estado
tf = 52;         % Tiempo final de simulación
h = 0.1;         % Tiempo de muestreo
L = tf/h;
%--- Generación de la señal medida
x(1) = 0;        % Estado inicial del sistema
y(1) = x(1) + R*randn(1); % Medición inicial
for i=1:L-1,
    x(i+1) = A*x(i) + Q*randn(1);
    y(i+1) = C*x(i) + R*randn(1);
end;
%--- Inicio de estimación de Kalman
xe(1) = 0;       % Estimación inicial del estado del sistema
P = 1;          % Matriz de covarianza inicial del error
for i=2:L,
    xm = A*xe(i-1);
    Pm = A*P*A' + Q;
    Pxy = Pm*C';
    Pyy = C*Pm*C' + R;
    GK = Pxy*Pyy^-1;
    ym = C*xm;
    xe(i) = xm + GK*( y(i) - ym );
    P = Pm - GK*C*Pm;
end;
```



La anterior Figura muestra un trayectoria estimada del filtro discreto de Kalman, mediante el algoritmo presentado.

### 7.2.3. Filtración óptima de sistemas continuos lineales

Además del cambio usual de las ecuaciones de diferencias a ecuaciones diferenciales, o de sumatorias a integrales, el modelo de filtración continuo, como generalización del discreto, es más complejo, en la medida en que la cantidad de observaciones en unidad de tiempo, en el caso continuo no es finita, sino infinita. En calidad de modelo de estudio, se tiene las siguientes ecuaciones de la planta y el dispositivo de medida (filtro de *Kalman-Bucy*):

$$\begin{cases} \dot{\mathbf{x}}(t) = \mathbf{A}(t)\mathbf{x}(t) + \mathbf{w}(t) \\ \mathbf{y}(t) = \mathbf{C}(t)\mathbf{x}(t) + \mathbf{v}(t) \end{cases}$$

para  $\forall t \geq t_0$ . El vector a estimar  $\mathbf{x}(t)$  es de dimensión  $n$ , para el cual se asume que  $\mathbf{x}(t_0)$  es un vector aleatorio con media cero y función de covarianza,

$$\mathbf{R}_{\mathbf{x}_0}(t_0) = E \left\{ (\mathbf{x}(t_0) - E\{\mathbf{x}(t_0)\}) (\mathbf{x}(t_0) - E\{\mathbf{x}(t_0)\})^\top \right\}$$

mientras, el vector de medida  $\mathbf{y}(t)$  tiene dimensión  $r$ . Las matrices  $\mathbf{A}$  y  $\mathbf{C}$  son conocidas. Para las perturbaciones  $\mathbf{w}(t)$ , con dimensión  $n$ , y los errores Gaussianos de medida, con dimensión  $r$ , se cumplen los siguientes momentos:

$$\begin{aligned} E\{\mathbf{w}(t)\} &= 0, E\{\mathbf{w}(t)\mathbf{w}(t)^\top\} = \mathbf{R}_w(t)\delta(t-\tau), \\ E\{\mathbf{v}(t)\} &= 0, E\{\mathbf{v}(t)\mathbf{v}(t)^\top\} = \mathbf{R}_v(t)\delta(t-\tau) \end{aligned}$$

Las matrices  $\mathbf{R}_w$  y  $\mathbf{R}_v$  son simétricas y con elementos variables en el tiempo. Las señales aleatorias  $\mathbf{w}(t)$  y  $\mathbf{v}(t)$  tienen función de correlación mutua cero. La estimación  $\hat{\mathbf{x}}(t)$  se debe hallar a partir de los resultados de medida  $\mathbf{y}(\tau)$ ,  $t_0 \leq \tau \leq t$ , que en el caso de la estimación lineal no sesgada, con el mínimo error cuadrático medio, se describe por la ecuación,

$$\frac{d}{dt}\hat{\mathbf{x}}(t) = \mathbf{A}(t)\hat{\mathbf{x}}(t) + \mathbf{K}(t)(\mathbf{y}(t) - \mathbf{C}(t)\hat{\mathbf{x}}(t)), \quad \hat{\mathbf{x}}(t_0) = 0,$$

La matriz de Kalman se calcula de la siguiente expresión,

$$\mathbf{K}(t) = \hat{\mathbf{R}}_{\mathbf{x}_0}(t) \mathbf{C}^\top(t) \mathbf{R}_v^{-1}(t) \quad (7.23)$$

La matriz de covarianza del error se determina por la *ecuación matricial diferencial de Riccati*

$$\begin{aligned} \frac{d}{dt}\hat{\mathbf{R}}_{\mathbf{x}_0}(t) &= \mathbf{A}(t)\hat{\mathbf{R}}_{\mathbf{x}_0}(t) + \hat{\mathbf{R}}_{\mathbf{x}_0}(t)\mathbf{A}^\top(t) - \hat{\mathbf{R}}_{\mathbf{x}_0}(t)\mathbf{C}^\top(t)\mathbf{R}_v^{-1}(t)\mathbf{C}(t)\hat{\mathbf{R}}_{\mathbf{x}_0}(t) + \mathbf{R}_w(t), \end{aligned} \quad (7.24)$$

asumiendo  $\hat{\mathbf{R}}_{\mathbf{x}_0}(t_0) = \mathbf{R}_{\mathbf{x}_0}(t_0)$ .

Si el modelo matemático está dado en la forma

$$\begin{cases} \dot{\mathbf{x}}(t) + \mathbf{A}(t)\mathbf{x}(t) + \mathbf{B}(t)\mathbf{u}(t) + \mathbf{w}(t), & E\{\mathbf{x}(t_0)\} = \xi \\ \mathbf{y}(t) = \mathbf{C}(t)\mathbf{x}(t) + \mathbf{v}(t) \end{cases}$$

entonces la estimación lineal no sesgada del vector de estados  $\mathbf{x}(t)$ , que cumple la condición del mínimo error cuadrático medio, se determina como la solución de la siguiente ecuación diferencial,

$$\frac{d}{dt}\tilde{\mathbf{x}}(t) = \mathbf{A}(t)\tilde{\mathbf{x}}(t) + \mathbf{B}(t)\mathbf{u}(t) + \mathbf{K}(t)(\mathbf{y}(t) - \mathbf{C}(t)\tilde{\mathbf{x}}(t)),$$

asumiendo

$$\mathbf{x}(t_0) = \xi$$

La matriz  $\mathbf{K}(t)$  se determina por la expresión (7.23), pero la matriz  $\mathbf{R}_{\mathbf{x}_0}(t)$  corresponde a la solución de la ecuación de Riccati en (7.24), la cual analíticamente tiene solución en muy pocos casos, por lo tanto, se considera su solución por métodos numéricos [49].

Uno de los casos, cuando la ecuación tiene solución, corresponde al siguiente. Sea la ecuación diferencial homogénea del error de filtración de la forma,

$$\dot{\tilde{\mathbf{x}}}(t) = (\mathbf{A}(t) - \hat{\mathbf{R}}_{\mathbf{x}_0}(t)\mathbf{C}^\top(t)\mathbf{R}_v^{-1}(t)\mathbf{C}(t))\tilde{\mathbf{x}}(t)$$

con su conjugado (si  $\dot{\tilde{\mathbf{x}}}(t) = \mathbf{F}(t)\tilde{\mathbf{x}}(t)$ , entonces la matriz conjugada será  $\dot{\xi}(t) = -\mathbf{F}^\top(t)\xi(t)$ ), de lo cual se obtiene,

$$\dot{\xi}(t) = (-\mathbf{A}^\top(t) + \mathbf{C}^\top(t)\mathbf{R}_v^{-1}(t)\mathbf{C}(t)\hat{\mathbf{R}}_{\mathbf{x}_0}(t))\xi(t)$$

La ecuación (7.24) se multiplica por  $\xi(t)$ , por lo que se obtiene,

$$\begin{aligned} & \frac{d}{dt}\hat{\mathbf{R}}_{\mathbf{x}_0}(t)\xi(t) \\ &= (\mathbf{A}(t)\hat{\mathbf{R}}_{\mathbf{x}_0}(t) + \hat{\mathbf{R}}_{\mathbf{x}_0}(t)\mathbf{A}^\top(t) - \hat{\mathbf{R}}_{\mathbf{x}_0}(t)\mathbf{C}^\top(t)\mathbf{R}_v^{-1}(t)\mathbf{C}(t)\hat{\mathbf{R}}_{\mathbf{x}_0}(t) + \mathbf{R}_w(t))\xi(t) \end{aligned}$$

luego, multiplicando la ecuación conjugada por  $\hat{\mathbf{R}}_{\mathbf{x}_0}(t)$  y reemplazando en la anterior ecuación, se tiene

$$\hat{\mathbf{R}}_{\mathbf{x}_0}(t)\dot{\xi}(t) + \hat{\mathbf{R}}_{\mathbf{x}_0}(t)\xi(t) = (\mathbf{A}(t)\hat{\mathbf{R}}_{\mathbf{x}_0}(t) + \mathbf{R}_w(t)\xi(t))$$

Como se observa, la última ecuación no presenta el término no lineal, entonces, al introducir la siguiente notación se obtiene un sistema lineal homogéneo del tipo,

$$\frac{d}{dt}\begin{bmatrix} \xi \\ \eta \end{bmatrix} = \begin{bmatrix} -\mathbf{A}^\top(t) & \mathbf{C}^\top(t)\mathbf{R}_v^{-1}(t)\mathbf{C}(t) \\ \mathbf{R}_w(t) & \mathbf{A}(t) \end{bmatrix}$$

donde los valores de  $\xi(t_0)$  y  $\eta(t_0)$  están dados a través de  $\mathbf{R}_{\mathbf{x}_0}(t)$ . Si se hallan los términos  $\xi(t)$  y  $\eta(t)$ , entonces al existir  $\xi^{-1}(t)$  se obtiene

$$\hat{\mathbf{R}}_{\mathbf{x}_0}(t) = \eta(t) \xi^{-1}(t)$$

En general, los conceptos de observabilidad y controlabilidad se pueden complementar de forma mutua, gracias al principio de dualidad de Kalman, según el cual la observación y la filtración del sistema inicial,

$$\begin{cases} \dot{\mathbf{x}}(t) &= \mathbf{A}(t) \mathbf{x}(t) + \mathbf{B}(t) \mathbf{u}(t), \\ \mathbf{y}(t) &= \mathbf{C}(t) \mathbf{x}(t), \quad t_0 \leq t \end{cases}$$

corresponden al control y regulación de un sistema dual, y viceversa.

El modelo matemático del sistema dual está dado por las expresiones

$$\begin{cases} -\dot{\xi}(t) &= \mathbf{A}^\top(t) \xi(t) + \mathbf{C}^\top(t) \mathbf{v}(t) \\ \eta(t) &= \mathbf{B}^\top(t) \xi(t) \end{cases}$$

de las cuales se puede demostrar que las propiedades del control óptimo permanecen vigentes, por lo menos desde el punto de vista algorítmico, en la síntesis de filtros óptimos [11].

En general, la tarea conjunta de estimación de los parámetros de la señal y de los estados del sistema es de carácter no lineal. Así sea el modelo analizado descrito por las relaciones,

$$\begin{cases} \dot{\mathbf{x}} &= \mathbf{f}_x \{\mathbf{x}, \mathbf{u}, \mathbf{a}, \xi, t\}; \\ \mathbf{y} &= \mathbf{f}_y \{\mathbf{x}, \mathbf{u}, \mathbf{C}, \eta, t\} \end{cases}$$

donde  $\mathbf{x}(t)$  es el vector de estados,  $\mathbf{u}(t)$  es la señal observada de entrada,  $\mathbf{a}(t)$  es el vector de parámetros de la planta,  $\xi(t)$  es el ruido de entrada o perturbación,  $\eta(t)$  es el error de medida,  $\mathbf{y}(t)$  es la señal observada de salida y  $\mathbf{C}$  es la matriz de observaciones.

El vector de parámetros de la planta  $\mathbf{a}$  se interpreta como la función dada por el control  $\dot{\mathbf{a}} = 0$  para un vector de condiciones iniciales  $\mathbf{a}(0)$ , luego, el modelo es,

$$\begin{bmatrix} \dot{\mathbf{x}} \\ \mathbf{a} \end{bmatrix} = \begin{bmatrix} \mathbf{f} \{\mathbf{x}, \mathbf{u}, \mathbf{a}, \xi, t\} \\ 0 \end{bmatrix}$$

que se puede transformar en la forma

$$\begin{bmatrix} \dot{\mathbf{x}} \\ \mathbf{a} \end{bmatrix} = \mathbf{M} \begin{bmatrix} \mathbf{x} \\ \mathbf{a} \end{bmatrix}$$

donde  $\mathbf{M}$  no depende ni de  $\mathbf{x}$  ni de  $\mathbf{a}$ . Entonces, inclusive para una planta lineal con dependencia lineal de parámetros, la tarea de estimación conjunta de los parámetros de la señal y de los estados del sistema tiene carácter no lineal con relación al vector de parámetros y al estado del sistema  $[\mathbf{x}^\top, \mathbf{a}^\top]$ .

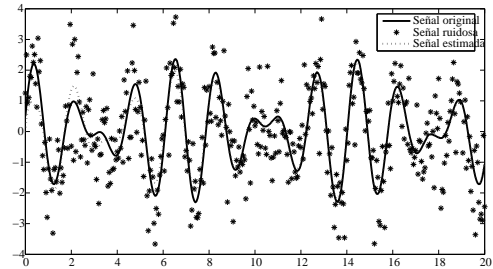
**Ejercicio en el CP 7.4.** Realiza un estimador, empleando la Filtración Kalman-Bucy, para un oscilador cuasiperiódico descrito por las siguientes ecuaciones diferenciales:

$$\begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \\ \dot{x}_3 \\ \dot{x}_4 \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ -\omega_1^2 & -2a & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & -\omega_2^2 & -2a \end{bmatrix} \begin{bmatrix} x_1(t) \\ x_2(t) \\ x_3(t) \\ x_4(t) \end{bmatrix} + \begin{bmatrix} 0 \\ 1 \\ 0 \\ 1 \end{bmatrix} w(t)$$

$$y(t) = \begin{bmatrix} 1 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} x_1(t) \\ x_2(t) \\ x_3(t) \\ x_4(t) \end{bmatrix} + v(t)$$

donde  $a = 0$ ,  $\omega_1 = 4$  y  $\omega_2 = \pi$ .

```
function KalmanBucy
% Oscilador armónico amortiguado
% Condiciones iniciales
xo = [0.5 4 0 4]'; xeo = zeros(4,1);
Po = 20*eye(4); r = 1;
ko=[xo;xeo;Po(:,1);Po(:,2);Po(:,3);Po(:,4)];
h = 5e-2; tspan = 0:h:20;
[t,k] = ode45(@KB,tspan,ko);
y = k(:,1)+k(:,3);
yn = y + r*randn(length(tspan),1);
ye = k(:,5)+k(:,7);
%-----
function dk = KB(t,k)
%-- Ajustes generales del filtro de Kalman
w1 = 4; w2 = pi; a = 0; r = 1; q = 0.1;
%-- Matrices del sistema
A = [ 0 1 0 0;
      -w1^2 -2*a 0 0;
      0 0 0 1;
      0 0 -w2^2 -2*a];
C = [1 0 1 0];
%-- Matriz de covarianza
Q = q*eye(4); Q(1,1) = 0; Q(3,3) = 0;
G = [0 1 0 1];
%-- Sistema real
dx = zeros(4,1); x = k(1:4); dx =
A*x + G*q*randn(4,1);
y = C*x + r*randn(1);
%-- Ecuación de Ricatti
P = [k(9:12) k(13:16) k(17:20) k(21:24)];
dP = zeros(4); K = P*C'*r^-1;
dP = A*P + P*A' - K*r*K' + Q;
%-- Estado estimado del sistema real
dxs = zeros(4,1); xs = k(5:8);
dxs = A*xs + K*(y-C*xs);
%-- Salida de datos
dk = zeros(20,1); dk(1:4) = dx;
dk(5:8) = dxs; dk(9:12) = dP(:,1);
dk(13:16) = dP(:,2); dk(17:20) = dP(:,3);
dk(21:24) = dP(:,4);
%--
```



### 7.2.4. Filtración de sistemas no lineales

Sea un sistema lineal en tiempo continuo descrito por la ecuación estocástica vectorial del tipo difusivo,

$$d\mathbf{x}(t) = \mathbf{f}\{\mathbf{x}, t\} dt + \mathbf{v}\{\mathbf{x}, t\} d\beta(t), \quad t \geq t_0 \quad (7.25)$$

en las cuales  $\beta(t)$  es un proceso estándar de Wiener, en el sentido en que los valores aleatorios son reales con distribución normal de incrementos, valor medio  $E\{d\beta\} = 0$  y varianza  $E\{d\beta^2\} = dt$ . Las funciones  $\mathbf{f}\{\cdot\}$  y  $\mathbf{v}\{\cdot\}$  son respectivamente los coeficientes de transporte y difusión.

El vector de estados del sistema  $\mathbf{x}(t)$  no puede ser observado directamente. Solamente es observable el vector  $\mathbf{y}(t)$ , por lo que el proceso de medida se describe también por otra ecuación diferencial estocástica,

$$d\mathbf{y}(t) = \mathbf{h}(\mathbf{x}, t) dt + d\gamma(t) \quad (7.26)$$

Los vectores  $\mathbf{x}$  y  $\mathbf{f}$  tienen dimensión  $n$ , mientras los vectores  $\mathbf{h}$  e  $\mathbf{y}$  son de dimensión  $r$ . Los procesos tipo Wiener  $\beta(t)$  y  $\gamma(t)$  se asumen con correlación mutua nula.

La estimación del vector de estados del sistema se puede realizar mediante la linealización de las ecuaciones (7.25) y (7.26), con el consecutivo empleo de métodos de filtración lineal, lo cual conlleva a las siguientes ecuaciones:

$$\begin{cases} \dot{\tilde{\mathbf{x}}}(t) &= \mathbf{f}\{\tilde{\mathbf{x}}, t\} + \mathbf{v}\{\tilde{\mathbf{x}}, t\} \beta(t) + \mathbf{R}_{\mathbf{x}} \left( \frac{\partial \mathbf{h}}{\partial \mathbf{x}} \right)^T \mathbf{R}_{\gamma}^{-1} (\mathbf{y}(t) - \mathbf{h}(\tilde{\mathbf{x}}, t)), \\ \dot{\mathbf{R}}_{\mathbf{x}} &= \frac{\partial \mathbf{f}}{\partial \mathbf{x}} \mathbf{R}_{\mathbf{x}_0} + \dot{\mathbf{R}}_{\mathbf{x}} \left( \frac{\partial \mathbf{f}}{\partial \mathbf{x}} \right)^T + \mathbf{v} \mathbf{R}_{\beta} \mathbf{v}^T - \mathbf{R}_{\mathbf{x}} \left( \frac{\partial \mathbf{h}}{\partial \mathbf{x}} \right)^T \mathbf{R}_{\gamma}^{-1} \frac{\partial \mathbf{h}}{\partial \mathbf{x}} \mathbf{R}_{\mathbf{x}} \end{cases}$$

con los siguientes momentos

$$\begin{aligned} E\{\mathbf{x}(t_0)\} &= \tilde{\mathbf{x}}(t_0) = \bar{\mathbf{x}}(t_0) \\ E\{(\mathbf{x}(t_0) - \bar{\mathbf{x}}(t_0))(\mathbf{x}(t_0) - \bar{\mathbf{x}}(t_0))^T\} &= \mathbf{R}_{\mathbf{x}}(t_0) = \mathbf{R}_{\mathbf{x}_0} \\ E\{(\dot{\beta}(t) - \bar{\dot{\beta}}(t))(\dot{\beta}(\tau) - \bar{\dot{\beta}}(\tau))^T\} &= \mathbf{R}_{\beta}(t) \delta(t - \tau) \\ E\{(\dot{\gamma}(t) - \bar{\dot{\gamma}}(t))(\dot{\gamma}(\tau) - \bar{\dot{\gamma}}(\tau))^T\} &= \mathbf{R}_{\gamma}(t) \delta(t - \tau) \\ E\{(\mathbf{x}(t_0) - \bar{\mathbf{x}}(t_0))(\dot{\beta}(\tau) - \bar{\dot{\beta}}(\tau))^T\} &= 0 \end{aligned}$$

siendo  $\bar{\mathbf{x}}(t_0)$ ,  $\bar{\dot{\gamma}}(t)$  y  $\bar{\dot{\beta}}(t)$  son los valores promedios de los respectivos procesos.

En el caso de análisis de la filtración de sistemas no lineales en tiempo discreto, se parte del proceso no lineal de dimensión múltiple, descrito en la forma,

$$\mathbf{x}[k+1] = \mathbf{f}^k\{\mathbf{x}[k]\} + \mathbf{w}[k], \quad k = 0, 1, \dots$$



para el cual, las medidas se realizan mediante un dispositivo, descrito por la ecuación no lineal,  $\mathbf{y}[k] = \mathbf{h}^k \{\mathbf{x}[k]\} + \mathbf{v}[k]$ ,  $k = 0, 1, \dots$

Las presunciones sobre la estructura de aleatoriedad de todas las variables estocásticas siguen siendo las mismas que se hicieron para el filtro Kalman-Bucy. El algoritmo de filtración no lineal se determina de las siguientes relaciones recurrentes:

$$\begin{aligned}\tilde{\mathbf{x}}[k] &= \bar{\mathbf{x}}[k] + \mathbf{K}[k] - \mathbf{h}^k \{\bar{\mathbf{x}}[k]\}, \quad \bar{\mathbf{x}}[k] = \mathbf{f}^k \{\tilde{\mathbf{x}}[k-1]\}, \\ \mathbf{R}_x[k] &= \frac{\partial \mathbf{f}^{k-1}}{\partial \mathbf{x}[k-1]} \hat{\mathbf{R}}[k-1] \left( \frac{\partial \mathbf{f}^{k-1}}{\partial \mathbf{x}[k-1]} \right)^\top + \mathbf{R}_w[k-1] \\ \mathbf{K}[k] &= \hat{\mathbf{R}}_x[k] \left( \frac{\partial \mathbf{h}^k}{\partial \mathbf{x}[k]} \right)^\top \left( \frac{\partial \mathbf{h}^k}{\partial \mathbf{x}[k]} \hat{\mathbf{R}}_x[k] \left( \frac{\partial \mathbf{h}^k}{\partial \mathbf{x}[k]} \right)^\top + \mathbf{R}_v[k] \right)^{-1} \\ \hat{\mathbf{R}}_x[k] &= \bar{\mathbf{R}}_x[k] - \mathbf{K}[k] \frac{\partial \mathbf{h}^k}{\partial \mathbf{x}[k]} \bar{\mathbf{R}}_x[k]\end{aligned}$$

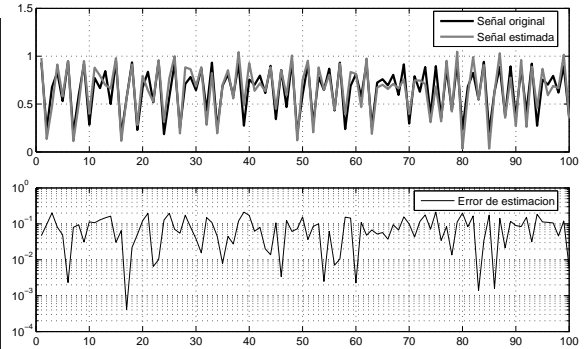
en las que se asume:  $\bar{\mathbf{R}}_x[0] = \mathbf{R}_x[0]$ , además, con condiciones iniciales conocidas,  $\tilde{\mathbf{x}}[0]$ .

El método de linealización es suficientemente efectivo, mientras no se tengan perturbaciones de potencia significativa, las que afectan la calidad de la estimación. Otro factor a tener en cuenta, en el rendimiento del filtro, está en la selección apropiada de los valores iniciales del vector estimación  $\tilde{\mathbf{x}}[0]$ .

**Ejercicio en el CP 7.5.** Realizar el filtro extendido de Kalman para la estimación de estado del mapa logístico, el cual se define mediante la siguiente ecuación de diferencias no lineal:

$$x[k] = \alpha(x[k-1] - x[k-1]^2)$$

```
%--Estimación del mapa logístico
function LMEstEKF
%-- Se genera la serie temporal
alpha = 3.8;
L = 5000;
x = zeros(L,1);
x(1) = 0.11;
for i=2:L,
    x(i) = alpha*x(i-1) - alpha*x(i-1)^2;
end;
y = x + 0.1*randn(L,1);
%-- Filtro extendido de Kalman
xe = zeros(L,1); xe(1) = 1;
P = 1e0;
Rw = 1e-10; Rv = 0.1;
for i=2:L,
    xe_ = alpha*xe(i-1) - alpha*xe(i-1)^2;
    A = alpha - 2*alpha*xe(i-1);
    Px_ = A*P*A' + Rw;
    K = Px_ * ( Px_ + Rv )^-1;
    xe(i) = xe_ + K*( y(i) - xe_ );
    Px = ( 1 - K )*Px_;
end;
%--
```



### 7.2.5. Filtración adaptativa de sistemas dinámicos

La filtración de Kalman, en variadas situaciones prácticas, puede ser no suficientemente efectivo, esto es, no brindar la suficiente precisión, o bien, simplemente no converger. Lo anterior es debido, entre otras, por las siguientes razones:

- Los parámetros en el modelo del sistema no concuerdan suficientemente con valores reales, durante su funcionamiento, por la dificultad en la estimación,
- La presunción de Gaussividad sobre la estructura, tanto en las perturbaciones como en los errores de medida, no es cierta en manera significativa.

Por lo tanto, es necesario mejorar el algoritmo de Kalman, de tal manera que se adapte a los factores a priori dados de no concordancia en el modelo, además a las diversas estructuras de aleatoriedad de las señales aleatorias durante el proceso.

Los métodos actuales para la resolución de esta tarea se orientan en dos aproximaciones básicas:

- La estimación, más bien la identificación, de los parámetros del modelo en el proceso de funcionamiento y el subsecuente empleo de las estimaciones obtenidas en el filtro convencional del Kalman.
- El análisis del carácter de evolución del sistema y los procesos aleatorios de medida y su adaptación por parte del filtro mediante cambios en su estructura de funcionamiento.

La segunda orientación, siendo más compleja en su implementación, se considera más efectiva, y es la que se analiza en este caso. Sea un sistema lineal en tiempo discreto dado en la forma,

$$\begin{cases} \mathbf{x}[k+1] &= \mathbf{A}[k] \mathbf{x}[k] + \mathbf{G}[k] \mathbf{w}[k] \\ \mathbf{y}[k] &= \mathbf{C}[k] \mathbf{x}[k] + \gamma[k] \alpha[k] + (1 - \gamma[k]) \mathbf{p}[k], \quad \mathbf{A}, \mathbf{C}, \mathbf{G} \in \mathbb{R}; \quad k = 0, 1, \dots \end{cases}$$

donde  $\mathbf{A}[k]$ ,  $\mathbf{C}[k]$  y  $\mathbf{G}[k]$  son matrices determinísticas de orden  $n \times n$ ,  $n \times r$  y  $n \times s$ , respectivamente,  $\mathbf{x}[k]$  es el vector de estados del sistema con dimensión  $n$ ,  $\mathbf{w}[k]$  son las perturbaciones aleatorias con dimensión  $s$ , independientes entre si, de estructura Gaussiana con valor medio cero y matriz de covarianza  $\mathbf{K}_{\mathbf{w}[k]}$ ,  $\mathbf{y}[k]$  es el vector de medidas con dimensión  $r$ .

Las magnitudes aleatorias independientes entre si, que forman la sucesión  $\gamma[k]$ , toman los valores 0 y 1 con probabilidades  $g_k$  y  $p_k$ :  $\mathbf{P}\{\gamma[k] = 1\} = p_k = 1 - g_k$ . Los vectores aleatorios  $\alpha[k]$  y  $\beta[k]$ , ambos de dimensión  $r$ , son Gaussianos con valor medio cero y función de covarianza  $\mathbf{K}_{\alpha[k]}$  y  $\mathbf{K}_{\beta[k]}$ , respectivamente. Generalmente, se supone que  $p_k \ll p_k$ , por lo que el modelo refleja la situación en la cual, en el canal de mediciones se presentan los ruidos  $\beta[k]$ , considerados típicos, y muy de vez en cuando aparecen interferencias de potencia fuerte  $\alpha[k]$ , que se consideran medidas anómalas.

El vector de valores iniciales  $\mathbf{x}[0]$  de estados del sistema, se asume Gaussiano con valor medio  $\mathbf{m}_{1\mathbf{x}[0]}$  y matriz de covarianza  $\mathbf{R}_{\mathbf{x}[0]}$ . Además, se supone que las sucesiones  $\{\mathbf{w}[k]\}$ ,  $\{\gamma[k]\}$ ,  $\{\alpha[k]\}$  y  $\{\beta[k]\}$  son mutuamente independientes, y a la vez, independientes del vector aleatorio  $\mathbf{x}[0]$ .

La adaptabilidad del algoritmo se basa en el principio de clasificación, como proceso de ordenamiento estadístico, de los resultados de observaciones  $\mathbf{y}[0]$  en dos grupos, tal que sean asociados a los ruidos de la clase  $\{\alpha[k]\}$  ó  $\{\beta[k]\}$ , con la consecuente modificación del algoritmo de filtración, a fin de tener en cuenta el modelo real de las perturbaciones. El procedimiento de clasificación se puede hacer, por ejemplo, empleando el método de toma de decisión bayesiano, mediante el cual la estimación de  $\gamma[k]$  se determina de la siguiente forma,

$$\hat{\gamma}[k] = \begin{cases} 0, & \Lambda_k \{\Delta \mathbf{y}[k]\} \geq \lambda_k \\ 1, & \Lambda_k \{\Delta \mathbf{y}[k]\} < \lambda_k \end{cases}$$

donde  $\Lambda_k \{\Delta \mathbf{y}[k]\}$  es la relación de verosimilitud, que para el caso se determina como

$$\Lambda_k \{\Delta \mathbf{y}[k]\} = \frac{p\{\Delta \mathbf{y}[k] | \gamma[k] = 1\}}{p\{\Delta \mathbf{y}[k] | \gamma[k] = 0\}}$$

siendo  $\lambda_k$  el umbral de decisión. En las anteriores expresiones,  $\Delta \mathbf{y}[k] = \mathbf{y}[k] - \mathbf{C}[k] \mathbf{m}_{1\mathbf{x}[k]}$ , en las cuales el valor medio del proceso  $\mathbf{x}[k]$  se puede determinar en forma recursiva, por ejemplo,  $\mathbf{m}_{1\mathbf{x}[k+1]} = \mathbf{A}[k] \mathbf{m}_{1\mathbf{x}[k]}$ ,  $k \geq 0$ .

La función densidad de probabilidad condicional del vector aleatorio  $\Delta \mathbf{y}[k]$  dado  $\gamma[k]$  está dada por la relación,

$$p\{\Delta \mathbf{y}[k] | \gamma[k]\} = \begin{cases} \sim \mathcal{N}\left(0, \hat{\mathbf{K}}_{\mathbf{x}[k]} - \mathbf{K}_{\alpha[k]}\right), & \gamma[k] = 1 \\ \sim \mathcal{N}\left(0, \hat{\mathbf{K}}_{\mathbf{x}[k]} - \mathbf{K}_{\beta[k]}\right), & \gamma[k] = 0 \end{cases}$$

donde  $\hat{\mathbf{K}}_{\mathbf{x}[k]} = \mathbf{C}[k] \mathbf{K}_{\mathbf{x}[k]} \mathbf{C}^T[k]$ .

De lo anterior, la relación de verosimilitud se calcula como

$$\Lambda_k \{\Delta \mathbf{y}[k]\} = \frac{|\mathbf{K}_{\mathbf{x}[k]+\beta[k]}|^{1/2}}{|\mathbf{K}_{\mathbf{x}[k]+\alpha[k]}|^{1/2}} \exp\left(\frac{1}{2} \Delta \mathbf{y}^T[k] \left(\mathbf{K}_{\mathbf{x}[k]+\beta[k]}^{-1} - \mathbf{K}_{\mathbf{x}[k]+\alpha[k]}^{-1}\right) \Delta \mathbf{y}[k]\right)$$

siendo  $\mathbf{K}_{\mathbf{x}[k]+\eta[k]}^{-1} = \hat{\mathbf{K}}_{\mathbf{x}[k]} + \mathbf{K}_{\eta[k]}$ ,  $\eta \in \{\alpha, \beta\}$ .

La función de covarianza se determina de la siguiente forma recursiva:

$$\hat{\mathbf{K}}_{\mathbf{x}[k+1]} = \mathbf{A}[k] \mathbf{K}_{\mathbf{x}[k]} \mathbf{A}^T[k] + \mathbf{G}[k] \mathbf{K}_{\mathbf{w}[k+1]} \mathbf{G}^T[k]$$

Como resultado, el algoritmo adaptativo para la obtención de solución estable  $\hat{\mathbf{x}}[k]$  en la tarea de filtración, toma la forma,

$$\begin{aligned}\hat{\mathbf{x}}[k] &= \mathbf{A}[k-1] \hat{\mathbf{x}}[k-1] \\ \bar{\mathbf{R}}_{\mathbf{x}}[k] &= \mathbf{A}[k-1] \hat{\mathbf{R}}_{\mathbf{x}}[k-1] \mathbf{A}^T[k-1] + \mathbf{G}[k-1] \mathbf{K}_{\mathbf{w}[k-1]} \mathbf{G}^T[k-1] \\ \tilde{\mathbf{x}}[k] &= \hat{\mathbf{x}}[k] + \mathbf{K}\{\tilde{\gamma}_k, k\} (\mathbf{y}[k] - \mathbf{C}[k] \hat{\mathbf{x}}[k])\end{aligned}$$

La ganancia del filtro adaptativo se determina de la forma ( $\forall k = 1, 2, \dots$ ),

$$\mathbf{K}\{\tilde{\gamma}_k, k\} = \bar{\mathbf{R}}_{\mathbf{x}}[k] \mathbf{C}^T[k] \left( \mathbf{C}[k] \bar{\mathbf{R}}_{\mathbf{x}}[k] \mathbf{C}^T[k] + \tilde{\gamma}[k] \mathbf{K}_{\alpha[k]} + (1 - \tilde{\gamma}[k]) \mathbf{K}_{\beta[k]} \right)^{-1}$$

donde  $\tilde{\gamma}_k = \{\tilde{\gamma}[i] : i = 0, \dots, k\}$ , siendo  $\tilde{\gamma}[i]$  la estimación del correspondiente valor de  $\gamma[i]$ . Los valores iniciales se dan como:  $\hat{\mathbf{x}}[0] = \bar{\mathbf{x}}[0] = \mathbf{m}_{1\mathbf{x}[0]}$ ,  $\bar{\mathbf{R}}_{\mathbf{x}}[0] = \mathbf{R}_{\mathbf{x}[0]}$ .

En esencia, el algoritmo adaptativo de Kalman presentado consiste de dos filtros, los cuales se conectan en concordancia con la decisión tomada sobre el modelo de perturbación presente. Sin embargo, ambos filtros tienen la misma estructura y difieren sólo en el valor de ganancia,  $\mathbf{K}\{\tilde{\gamma}_k, k\}$ , que depende del tipo de perturbación asumida.

El umbral de decisión  $\lambda_k$ , generalmente, se fija de manera experimental para cada aplicación concreta, por ejemplo mediante un modelo de imitación. Aunque existen aproximaciones que intentan fijar este valor de manera automática, así en [50], se presenta un método basado en el análisis discriminante óptimo bayesiano, para una matriz dada de pérdida.

De otra parte, en el algoritmo de Kalman es básica la información que brindan las respectivas funciones de covarianza de los perturbaciones y señales, las cuales no siempre se disponen, por lo tanto, el algoritmo debe tener en cuenta, alguna forma de su estimación. El siguiente algoritmo estima los valores de las funciones de covarianza.

Sea un sistema dinámico lineal en tiempo discreto, descrito por las ecuaciones recurrentes,

$$\begin{cases} \mathbf{x}[k+1] &= \mathbf{A}\mathbf{x}[k] + \mathbf{G}\mathbf{w}[k], \\ \mathbf{y}[k] &= \mathbf{C}\mathbf{x}[k] + \mathbf{v}[k], \quad k = 0, 1, \dots \end{cases} \quad (7.27)$$

donde  $\mathbf{x}[k+1]$  es de estados del sistema con dimensión  $n$ ,  $\mathbf{A}$  es la matriz de transición de orden  $n \times n$ ,  $\mathbf{C}$  y  $\mathbf{G}$  son matrices de orden  $r \times n$  y  $n \times s$ , respectivamente. Se asume que los vectores aleatorios  $\mathbf{w}[k]$  y  $\mathbf{v}[k]$ , con dimensiones respectivas  $s$  y  $r$ , son procesos Gaussianos con correlación nula, y para los cuales se definen los siguientes momentos,

$$\begin{aligned}E\{\mathbf{w}[k]\} &= 0, \quad E\{\mathbf{w}[k] \mathbf{w}[m]^T\} = \mathbf{R}_{\mathbf{w}} \delta_{km} \\ E\{\mathbf{v}[k]\} &= 0, \quad E\{\mathbf{v}[k] \mathbf{v}[m]^T\} = \mathbf{R}_{\mathbf{v}} \delta_{km} \\ E\{\mathbf{v}[k] \mathbf{w}[m]^T\} &= 0\end{aligned}$$

Las matrices de covarianza consideradas,  $\mathbf{R}_{\mathbf{w}}$  y  $\mathbf{R}_{\mathbf{v}}$ , son positivas semidefinidas. Además, se asume que el vector inicial de estados del sistema es Gaussiano con valor medio 0 y matriz de covarianza  $\mathbf{R}_{\mathbf{x}_0}$ . Por último, se asume que el sistema, (7.27), es observable.

De manera similar, al caso del algoritmo adaptativo anterior, la estimación se basa en el análisis de las propiedades de restablecimiento en el tiempo del proceso de filtración, esto es, de la sucesión

$$\Delta \mathbf{y}[k] = \mathbf{y}[k] - \mathbf{C}\hat{\mathbf{x}}[k]$$

El proceso, en su restablecimiento, contiene la nueva información estadística obtenida mediante las observaciones  $\mathbf{y}[k]$ , y la cual sirve para determinar la estimación las respectivas matrices de covarianza  $\mathbf{R}_w$  y  $\mathbf{R}_v$  [11].

Cuando el sistema llega alcanza el estado estable, tanto el coeficiente de ganancia del filtro  $\mathbf{K}$ , como la matriz de covarianza del error extrapolada  $\hat{\mathbf{K}}$  no cambian en el tiempo,

$$\mathbf{K} = \hat{\mathbf{R}}_x \mathbf{C}^\top (\mathbf{C} \hat{\mathbf{R}}_x \mathbf{C}^\top + \mathbf{R}_v)^{-1} \quad (7.28a)$$

$$\hat{\mathbf{R}}_x = \mathbf{A} (\mathbf{I} - \mathbf{K} \mathbf{C}) \hat{\mathbf{R}}_x (\mathbf{I} - \mathbf{K} \mathbf{C})^\top + \mathbf{A} \mathbf{K} \mathbf{R}_v \mathbf{K}^\top \mathbf{A}^\top + \mathbf{G} \mathbf{R}_w \mathbf{G}^\top \quad (7.28b)$$

El par de ecuaciones (7.28a) y (7.28b) sirven para el cálculo de la matriz  $\mathbf{R}_x$ , bien en la estimación óptima, bien en el empleo de algoritmos subóptimos que estiman las matrices de covarianza  $\mathbf{R}_w$  y  $\mathbf{R}_v$ .

Cabe anotar, que la relación estadística entre diferentes valores de observación está dada por su momento conjunto,  $\mathbf{C}_k = E \{ \Delta \mathbf{y}[k] \Delta \mathbf{y}^\top[i - k] \}$ . De otra parte, empleando la definición del proceso de restablecimiento, se pueden obtener las siguientes relaciones importantes en la estimación de las funciones de covarianza buscadas:

$$\begin{aligned} \mathbf{C}_0 &= \mathbf{C} \hat{\mathbf{R}}_x \mathbf{C}^\top + \mathbf{R}_v, \\ \mathbf{C}_k &= \mathbf{C} (\mathbf{A} (\mathbf{I} - \mathbf{K} \mathbf{C}))^{k-1} \mathbf{A} (\hat{\mathbf{R}}_x \mathbf{C}^\top - \mathbf{K} \mathbf{C}_0), \quad k \geq 1 \end{aligned}$$

Básicamente, el algoritmo presente de filtración adaptativa consiste, en que una vez dadas en el primer momento, las estimaciones a priori  $\mathbf{R}_{w[0]}$  y  $\mathbf{R}_{v[0]}$  de las respectivas funciones de covarianza, se realiza el perfeccionamiento de su estimación en el proceso de filtración.

El procedimiento de estimación de las matrices de covarianza es el siguiente:

- Estimación de la matriz  $\mathbf{y}_k$  a partir de la definición

$$\tilde{\mathbf{C}}_k = \frac{1}{N} \sum_{i=k}^N \Delta \mathbf{y}[k] \Delta \mathbf{y}^\top[i - k]$$

siendo  $N - k$  la cantidad de observaciones en análisis.

- Estimación del primer término de multiplicación en (7.28a). Por cuanto,

$$\begin{aligned} \mathbf{C}_1 &= \mathbf{C} \hat{\mathbf{R}}_x \mathbf{C}^\top - \mathbf{C} \mathbf{A} \mathbf{K} \mathbf{C}_0 \\ &\dots \quad \dots \quad \dots \\ \mathbf{C}_n &= \mathbf{C} \mathbf{A}^n \hat{\mathbf{R}}_x \mathbf{C}^\top - \mathbf{C} \mathbf{A} \mathbf{K} \mathbf{C}_{n-1} - \dots - \mathbf{C} \mathbf{A}^n \mathbf{K} \mathbf{C}_0 \end{aligned}$$

entonces

$$\hat{\mathbf{R}}_{\mathbf{x}} \mathbf{C}^T = \mathbf{B} + \begin{bmatrix} \mathbf{C}_1 + \mathbf{CAK}\mathbf{C}_0 \\ \mathbf{C}_2 + \mathbf{CAK}\mathbf{C}_1 + \mathbf{CA}^2\mathbf{K}\mathbf{C}_0 \\ \vdots \\ \mathbf{C}_n + \mathbf{CAK}\mathbf{C}_{n-1} + \cdots + \mathbf{CA}^n\mathbf{K}\mathbf{C}_0 \end{bmatrix}$$

siendo  $\mathbf{B}^+$  la respectiva matriz pseudotranspuesta (de orden  $1 \times n$ )

$$\mathbf{B} = \begin{bmatrix} \mathbf{C} \\ \mathbf{CA} \\ \vdots \\ \mathbf{CA}^{n-1} \end{bmatrix}$$

que corresponde a la multiplicación de las matrices de observabilidad y de transición  $\mathbf{A}$ . Entonces, en calidad de estimación de la matriz  $\mathbf{PC}^T$  se emplea la expresión,

$$\hat{\mathbf{R}}_{\mathbf{x}} \tilde{\mathbf{C}}^T = \mathbf{B}^+ \begin{bmatrix} \tilde{\mathbf{C}}_1 + \mathbf{CAK}\mathbf{C}_0 \\ \tilde{\mathbf{C}}_2 + \mathbf{CAK}\tilde{\mathbf{C}}_1 + \mathbf{CA}^2\mathbf{K}\mathbf{C}_0 \\ \vdots \\ \tilde{\mathbf{C}}_n + \mathbf{CAK}\tilde{\mathbf{C}}_{n-1} + \cdots + \mathbf{CA}^n\mathbf{K}\mathbf{C}_0 \end{bmatrix}$$

- Cálculo de la estimación de la función de covarianza  $\mathbf{R}_{\mathbf{v}}$ , empleando la siguiente expresión  $\tilde{\mathbf{R}}_{\mathbf{v}} = \mathbf{K}\tilde{\mathbf{C}}_0 - \mathbf{C}(\hat{\mathbf{R}}_{\mathbf{x}}\mathbf{C}^T)$
- Cálculo de la estimación de la función de covarianza  $\mathbf{R}_{\mathbf{w}}$ , mediante la solución del siguiente sistema,

$$\begin{aligned} \sum_{i=0}^{k-1} \mathbf{CA}^i \mathbf{G} \tilde{\mathbf{R}}_{\mathbf{w}} \mathbf{G}^T (\mathbf{A}^{i-k})^T \mathbf{C}^T \\ = (\hat{\mathbf{R}}\mathbf{C}^T)^T (\mathbf{A}^{-k}) \mathbf{C}^T - \mathbf{CA}^k \hat{\mathbf{R}}\mathbf{C}^T - \sum_{i=0}^{k-1} \mathbf{CA}^i \tilde{\mathbf{S}} (\mathbf{A}^{i-k})^T \mathbf{C}^T \end{aligned}$$

siendo  $\tilde{\mathbf{S}} = \mathbf{A} \left( -\mathbf{K} (\hat{\mathbf{R}}\mathbf{C}^T)^T - \hat{\mathbf{R}}\mathbf{C}^T \mathbf{K}^T + \mathbf{KC}_0\mathbf{K}^T \right) \mathbf{A}^T$ .

La estimación  $\tilde{\mathbf{C}}_k$ , para de disminuir el coste computacional, se realiza de forma recursiva,

$$\tilde{\mathbf{C}}_k^{N+1} = \tilde{\mathbf{C}}_k^N + \frac{1}{N} \left( \Delta \mathbf{y}[n+1] \Delta \mathbf{y}^T[n+1-k] - \tilde{\mathbf{C}}_k^N \right)$$

donde  $\tilde{\mathbf{C}}_k^l$  es la estimación de  $\mathbf{C}_k$  sobre  $l$  observaciones.

# Bibliografía

- [1] I. I. Privalov, *Vvedenie v teoriju funktsii kompleksnogo peremennogo*. Moskva: Nauka, 1977.
- [2] I. S. Gonorovski, *Circuitos y señales radiotécnicos*. Moscú: Progress, 1972.
- [3] L. A. Zalmanson, *Fourier, Haar, Walsh Transforms and their application in Control, Communications and other Fields*. Moscow: Nauka, 1989.
- [4] L. E. Franks, *Signal Theory*. New Jersey: PH. Englewood Cliffs, 1969.
- [5] R. E. Edwards, *Fourier Series: A Modern Introduction*. Camberra: SV, 1979, vol. I, II.
- [6] G. V. Obrezkov, *Prikladnye Matematicheskie Metody Analiza v Radiotekhnike*. Moskva: VysshajaShkola, 1985.
- [7] N. Wiener, *Nonlinear Problems in Random Theory*, 1st ed. Cambridge, Mass: MIT Press, 1958.
- [8] L. V. Danilov, *Rjady Volterra-Picard v teorii njelinejnyj elektricheskij tsepjey*. Moskva: Radio i Svjaz, 1987.
- [9] V. I. Tijonov, *Nielinijnye prjeobrazovaniya sluchaynyj protsessov*. Moskva: Radio i Svjaz, 1986.
- [10] V. D. Ender and N. A. Verhoeckx, *Discrete-Time Signal Processing. An Introduction*. Cambridge: PHI, 1989.
- [11] V. Bolnokin and P. Chinaev, *Analiz i Sintez Sistem Avtomaticheskogo Upravljenja v EBM*. Moskva: Radio i Svjaz, 1991.
- [12] B. R. Levin, *Teoritieskije osnovy statisticheskoy radiotekhniki*. Moskva: Radio i Svjaz, 1989.
- [13] A. Zaezdnyi, *Osnivy Raschetov po Statisticheskoy Radiotekhnike*. Moskva: Svjaz, 1969.
- [14] G. R. Cooper and C. D. McGuillen, *Probabilistic Methods of Signal and System Analysis*. Pardue University, 1986.
- [15] V. I. Tijonov, *Statisticheskaja radiotekhnika*. Moskva: Radio i Svjaz, 1982.
- [16] A. Lebedev and E. Chernyakovsky, *Verojatnosnije metody v vycheslitel'noy tejnike*. Moskva: Vysshaya Shkola, 1986.
- [17] V. S. Korolyuk and N. Portenko, *Spravochnik por teorii verojatnostey i matematicheskoy statistike*. Moskva: Nauka, 1978.
- [18] A. A. Borovkov, *Estadística Matemática: Estimación, verificación de hipótesis*. Moscú: Mir, 1988.
- [19] E. Lvovski, *Statisticheskije metody postrojenija empiricheskij formul*. Moskva: Vysshaja Shkola, 1988.
- [20] W. Lederman, *Handbook of applicable mathematics*, E. Lloyd, Ed. New York: John Wiley Sons, 1984, vol. VI:Statistics.
- [21] S. Yu, *Theory of the Detection and Evaluation of Stochastic Signals (in Russian)*. Moscow: Sovetskoe Radio, 1978.

- [22] P. S. Akimov, A. I. Senin, and V. I. Solenov, *Signaly i ij obrabotkav informatsionnyj sitemaj*. Moskva: Radio i Svjaz, 1994.
- [23] S. Vilenkin, *Statisticheskiye metody issledovaniya sistem avtomaticheskogo regulirovaniya*. Moskva: Sovetskoe Radio, 1967.
- [24] D. Hilbert and R. Kurant, *Métody matematicheskoi fiziki*. Moskva: Gostechizdat, 1951, vol. T.1.
- [25] B. Shajtarin and V. Kovrigin, *Metody spektralnogo otsenivaniya sluchaiynyj proytssessov*. Moskva: Geliyus ARV, 2005.
- [26] J. Bendat and A. Piersol, *Random data: analysis and measurement procedures*. New York, USA: John Wiley and Sons, 1971.
- [27] L. R. Rabiner and B. Gold, *Theory and application of digital signal processing*. Prentice Hall, 1975.
- [28] L. Goldemberg and B. Matiyushkin, *Tsifrovaja Obrabotka signalov*. Moskva: Radio i Svjaz, 1985.
- [29] L. Marple, *Digital Spectral Anaylisis with Applications*. Englewood Cliffs, New Jersey: Prentice Hall, 1987.
- [30] V. I. Tijonov and M. A. Mironov, *Markovskie protseccy*, moskva ed. SovetskoeRadio, 1977.
- [31] R. L. Stratanovich, *Uslovnye Markovskie Protsessy i ij primininje k teorii optimalnogo upravljenja*. Moskva: MGU, 1966.
- [32] Y. I. Paraev, *Vvedenije v statisticheskiju dinamiku protsessov upravljenja i filtratsii*, moskva ed. Sovetskoe Radio, 1976.
- [33] H. Xuedong, A. Acero, and H. Hsiao-wuen, *Spoken Language Processing*. New Jersey: Prentice Hall, 2001.
- [34] A. P. Dempster and N. Laird, "Maximum likelihood from incomplete data via the EM algorithm (with discussion)," *Journal of The Royal Statistica Society B*, vol. 39, pp. 1–38, 1977.
- [35] A. Ljolje, Y. Ephraim, and L. Rabiner, "Estimation of hidden markov models parameters by minimizing empirical error rate," in *Proceedings of International Conference on Acoustics, Speech and Signal Processing*, 1990, pp. 709–712.
- [36] S. E. Levinson, "Continuously variable duration hidden markov models for speech analysis," in *Proceedings of the International Conference in Acoustics, Speech and Signal Processing*, 11, Ed., 1986, pp. 1241–1244.
- [37] L. Rabiner, "A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition," *Proceedings of The IEEE*, vol. 77, no. 2, Febrero 1989.
- [38] M. Do, "Fast approximation of kullback-leibler distance for dependence trees and hidden markov models," *IEEE Signal Processing Letters*, vol. 10, no. 3, pp. 115–118, 2003.
- [39] R. Lyngso and C. Pedersen, "Metrics and similarity measures for hidden Markov models," in *Proceedings of The Seventh International Conference on Intelligent Systems for Molecular Biology*, 1999, pp. 178–186.
- [40] M.Ñechyba and Y. Xu, "Stochastic Similarity for Validating Human Control Strategy Models," *IEEE Transactions On Robotics And Automation*, vol. 14, no. 3, Junio 1998.
- [41] V. I. Tijonov and N. K. Kulman, *Njelinjeynaja filtratsia i kvasikogerentnyi prijem signalov*. Moskva: Sovetskoe Radio, 1975.
- [42] V. I. Tijonov and V.Ñ. Harisov, *Statistichesii Analiz i sintez radiotejnicheskij ustroystv i sistem*. Moskva: Radio i Svjaz, 2004.
- [43] W. B. Davenport and W. L. Root, *An Introduction to the Theory of Random Signals and Noise*. New York: McGraw-Hill, 1958.
- [44] I. S. Guelfand and A. A. Jaglom, *Integrirovanie v Funktsionalnyj Prostranstvaj*. Moskva: YspejMat-Nauk, 1951.



- 
- [45] E. Aved'jan and J. Tzypkin, "Obobschennyi algoritm kachmazha," *Avtomatika i Telemekhanika*, vol. 1, pp. 72–78, 1979.
  - [46] J. Markl, "Uskorenije sjodimosti algoritma kachmazha v sluchaje vremennoi korrelatsii vjednogo protsesssa," *Avtomatika i Telemekhanika*, vol. 8, pp. 70–73, 1980.
  - [47] P. Eikhoff, *System Identification*. Englewoods Cliffs,NJ: Prentice-Hall, 1975.
  - [48] R. Liptzer and A. Shirjaev, *Statistika Sluchainykh Protsessov*. Moskva: Nauka, 1974.
  - [49] M. Ageev, V. Alik, and J. Markov, *Biblioteka Algoritmov 101b-150b*. Moskva: Sovetskoe Radio, 1978.
  - [50] A. Ershov, "Stabil'nye metody otsenki parametrov," *Avtomatika i Telemekhanika*, vol. 8, pp. 66–100, 1978.



# Índice alfabético

- algoritmo
  - máxima esperanza, 206
  - propagación, 204
  - viterbi, 205
- Bayes
  - decisión, 110
  - estimación, 93
  - teorema, 44
- cadena
  - compuesta, 179
  - homogénea, 180
- Chebyshev
  - desigualdad, 47
- coeficiente
  - asimetría, 50
  - exceso, 50
- convolución
  - integral, 10
- correlación
  - estimación, 145
  - función, 69, 117, 124
  - intervalo, 125
  - matriz, 117
  - mutua, 231
- correlograma
  - definición, 161
- costos
  - función, 267
- covarianza
  - función, 69
  - matriz, 77
- decisión
  - óptima, 105
- densidad
  - espectral, 7
- DEP
  - definición, 132
  - estimación, 150
- detección
  - binaria, 268
- dicretización
  - teorema, 25
- Dirichlet
  - condiciones, 8
- distancia, 2
  - Kullback - Leibler, 221
- distribución
  - átomo, 45
- Ecuación
  - aproximación cero, 237
- ecuación
  - Chapman-Kolmogorov, 183
  - Fokker-Plank-Kolmogorov, 197
  - Ito, 194
  - Kolmogorov, 184
  - Kolmogorov-Feller, 196
  - Markov, 188
  - Ricatti, 312
  - Stratanovich, 193, 258
- ensamble, 115
- entropía
  - definición, 80
- ergodicidad
  - definición, 121, 122
- error
  - potencia, 85
- espacio
  - definición, 1
  - probabilístico, 42
- espectro
  - definición, 3
- estacionariedad
  - definición, 116
- estado
  - duración, 217
  - ecuación, 37
  - estable, 19
  - matriz, 37
  - sistema, 31
  - variables, 31, 296
  - vector, 33
- estimación

- definición, 85
- propiedades, 143
- puntual, 86
- seudobayesiana, 113
- FDP
  - definición, 46
  - Fisher, 60
  - Gamma, 58
  - Gauss, 55
  - Poisson, 62
  - Student, 59
  - transformación, 72
- filtración
  - Discreta, 27
- filtro
  - acoplado, 290
  - activo, 293
  - formante, 171
  - pasivo, 293
- Fisher
  - matriz, 278
- Fourier
  - $\mathcal{S}$ -espectro, 5
  - serie, 3
  - transformada, 7
- función
  - base, 2
  - característica, 123
  - generatriz, 68
  - ortogonal, 3
  - probabilidad, 53
  - transferencia, 15, 27
- Hilbert
  - espacio, 49
- hipótesis
  - prueba, 96
- independencia
  - estadística, 43
- información
  - medidas, 78
- intervalo
  - confianza, 91
- Ito
  - integral, 194
- Jensen
  - desigualdad, 209
- K-L
  - transformada, 129
- Kachmazh
  - método, 303
- Kalman
  - filtración, 308
- Kalman-Bucy
  - filtro, 312
- Kolmogorov
  - axiomas, 42
- Kotelnikov
  - teorema, 25
- Laplace
  - integral, 58
  - transformada, 10
- mínimos cuadrados
  - modelo, 99
- MacLaurin
  - serie, 51
- Markov
  - cadena, 178
  - modelos ocultos, 201
  - procesos, 127, 177
  - sucesiones, 198
  - transformación, 242, 249
- Minimax
  - detección, 271
  - Neymann-Pearson, 271
- modelo
  - AR, 172
  - ARMA, 172
  - MA, 172
  - paramétrico, 171
- momentos
  - centralizados, 47
  - cumulantes, 51
  - estimación, 144
  - iniciales, 46
  - relación, 52
- Neyman-Pearson
  - criterio, 97
- norma, 2
- operador
  - lineal, 12, 14, 234
  - método, 16
- pérdida
  - función, 105
- periodograma
  - algoritmo, 164
  - Bartlett, 158
  - Blackman-Tukey, 162
  - definición, 152
  - método, 157
  - Welch, 160
- promedios
  - tiempo, 116
- RBG
  - definición, 136
- redundancia
  - definición, 82
- regresión
  - estimación, 153
  - lineal, 100
  - modelo, 98
- respuesta

- a impulso, 15, 16, 27
- riesgo
  - función, 107
- señal
  - aleatoria, 115
- spline
  - definición, 99
- Stieltjes
  - integral, 141
- transformación
  - lineal, 12
- transformada
  - discreta, 26
- trayectoria, 115
- valores
  - medios, 46
- varianza
  - propiedades, 48
- ventana
  - espectral, 166
- verosimilitud
  - función, 88
- Volterra
  - series, 21, 243
- Wiener
  - filtración, 307
  - método, 20
- Wiener-Hopf
  - ecuación, 155, 284
- Wiener-Jinchin
  - transformada, 134, 232
- Yule-Walker
  - ecuaciones, 174



# Introducción

Several years ago I reached the conclusion that the Theory of Probability should no longer be treated as adjunct to statistics or noise or any terminal topic, but should be included in the basic training of all engineers and physicists as a separate course

---

Papoulis

El interés por el estudio de las señales aleatorias<sup>1</sup> sigue siendo actual, aún cuando su aplicación asociada a la tecnología e ingeniería cuenta con varias décadas de desarrollo. Particular importancia tienen los métodos y técnicas de análisis en presencia de aleatoriedad en señales y sistemas, que estén relacionados con las tecnologías de proceso discreto, en las cuales el empleo de métodos efectivos matemáticos se ha hecho masivo, gracias a la aparición de dispositivos digitales de alto rendimiento.

El estudio del fenómeno de aleatoriedad en señales y sistemas está directamente relacionado con el análisis de las propiedades de interacción cuando se tiene algún grado de incertidumbre o información incompleta. La principal fuente de aleatoriedad está asociada, entre otros factores, con las perturbaciones de medida, la presencia de señales extrañas, y tal vez el más importante, la información útil que llevan las mismas señales. En este sentido, es cada vez mayor el interés por el empleo de métodos de proceso digital de señales, basados en matemática estadística, que se orientan principalmente a la solución de dos tareas: la representación adecuada del conjunto de señales para la solución de un problema concreto, que tenga en cuenta las condiciones reales de la aplicación y, segundo, el proceso óptimo de las señales recogidas.

El material dispuesto en el presente texto describe las formas básicas de representación y proceso de señales aleatorias, con especial énfasis en los modelos de análisis matemático estadístico, que se considera, tienen un aporte significativo en la formulación y solución de aplicaciones en las áreas de sismología, análisis de bioseñales, sistemas de medida, sistemas de control y seguimiento, radiocomunicación, entre otros. Por eso, el contenido presentado corresponde a la evolución en el análisis de aleatoriedad desde las señales hasta su asociación con los respectivos sistemas de proceso. En cada sección se presentan, tanto ejemplos como problemas (incluyendo algunos ejercicios en el computador), que tienen como intención mejorar y complementar la percepción del material teórico analizado.

El contenido del texto es el siguiente: el capítulo I describe los fundamentos de representación determinística de señales y sistemas. El capítulo II presenta la caracterización de variables aleatorias y las técnicas básicas de estimación. Aunque el material de ambos capítulos se supone conocido, por ejemplo de los cursos de *Teoría de Señales* y *Teoría de Probabilidades*, es pertinente su inclusión a fin de recordar aspectos y definiciones importantes que se hacen necesarios para el entendimiento de los capítulos posteriores. El capítulo III describe las particularidades en la representación de procesos aleatorios en el tiempo, incluyendo su análisis experimental. Debido a la importancia de los procesos estocásticos de Markov, estos se analizan por a parte en el capítulo IV, con especial atención en el análisis experimental mediante los modelos ocultos de Markov. El capítulo V corresponde al análisis de la transformación de señales aleatorias en sistemas, así como la formación de procesos aleatorios mediante sistemas lineales. El capítulo VI presenta los principales métodos de detección y filtración de señales aleatorias, entendida como la optimización de la forma de representación de los dispositivos de proceso.

Por último, la necesidad de aumentar la efectividad del funcionamiento de los sistemas de control, en condiciones de factores aleatorios, ha estimulado el desarrollo de métodos de optimización orientada a

---

<sup>1</sup> *Aleatorio*. Del Latín *aleatorius*, propio del juego de dados. Real Academia de la Lengua

mejorar el registro y precisión de la información sobre las propiedades de la planta a diseñar, así como de los regímenes de su funcionamiento. En este sentido, el capítulo VII describe el análisis de sistemas estocásticos, en particular, se analiza la filtración Kalman.

El material está orientado a los estudiantes de posgrado, que requieran profundizar en el análisis de procesos aleatorios. Además, se puede emplear como material de ayuda en el curso de *Procesos Estocásticos*. El material teórico presentado tiene carácter de referencia, y por esto no se da la deducción de varias de las expresiones, brindándose la literatura necesaria para la profundización de cada tema en particular.

Finalmente, se agradece a los profesores Julio F. Suárez, Gerard Olivar y Sergey Adzhemov por sus consejos y correcciones hechos para dar mayor precisión y entendimiento del presente trabajo. Se aprecia también la colaboración que prestaron los estudiantes del programa posgrado, Jorge Jaramillo, Luis G. Sánchez, Julián D. Arias, Mauricio Álvarez, David Avendaño, Andrés F. Quiceno y todos aquellos, quienes pacientemente leyeron, siguieron y colaboraron con la evolución de los diferentes borradores del texto.

---

*Los Autores, 2007*



# Índice general

<b>1. Señales y sistemas</b>	<b>1</b>
1.1. Representación discreta de señales . . . . .	1
1.1.1. Espacio de representación de señales . . . . .	1
1.1.2. Descomposición en funciones ortogonales . . . . .	3
1.1.3. Ejemplos de conjuntos ortogonales completos . . . . .	3
1.2. Representación integral de señales y sistemas . . . . .	7
1.2.1. Densidad espectral . . . . .	7
1.2.2. Transformaciones lineales . . . . .	11
1.2.3. Representación integral de sistemas lineales . . . . .	13
1.2.4. Representación de sistemas no lineales . . . . .	18
1.3. Discretización de señales y sistemas . . . . .	25
1.3.1. Discretización uniforme . . . . .	25
1.3.2. Transformadas ortogonales discretas . . . . .	26
1.3.3. Representación ortogonal de sistemas en tiempo discreto . . . . .	27
1.4. Representación dinámica de sistemas . . . . .	31
1.4.1. Sistemas lineales en variables de estado . . . . .	31
1.4.2. Solución de la ecuación de estado . . . . .	37
<b>2. Variables aleatorias</b>	<b>41</b>
2.1. Valores y funciones de probabilidad . . . . .	41
2.1.1. Espacio probabilístico de variables . . . . .	41
2.1.2. Valores medios y momentos . . . . .	46
2.1.3. Funciones de probabilidad . . . . .	53
2.1.4. Modelos de funciones de probabilidad con dimensión múltiple . . . . .	67
2.1.5. Medidas de información en variables aleatorias . . . . .	78
2.2. Estimación en variables aleatorias . . . . .	85
2.2.1. Estimación puntual de momentos . . . . .	86
2.2.2. Intervalos de confianza . . . . .	91
2.2.3. Estimación de parámetros en la distribución normal . . . . .	94
2.2.4. Prueba de hipótesis . . . . .	96
2.2.5. Estimación de dependencias funcionales . . . . .	98
2.3. Teoría de decisión estadística . . . . .	105
2.3.1. Definiciones básicas . . . . .	105
2.3.2. Decisión Bayesiana . . . . .	110
2.3.3. Decisión no Bayesiana . . . . .	112
<b>3. Señales aleatorias</b>	<b>115</b>
3.1. Señales aleatorias en el tiempo . . . . .	115

3.1.1.	Estacionariedad de las señales aleatorias . . . . .	116
3.1.2.	Ergodicidad de las señales aleatorias . . . . .	121
3.1.3.	Descomposición espectral de señales aleatorias . . . . .	127
3.1.4.	Densidad espectral de potencia . . . . .	130
3.1.5.	Convergencia y continuidad de procesos aleatorios . . . . .	138
3.2.	Análisis experimental de señales estacionarias . . . . .	143
3.2.1.	Estimación de momentos . . . . .	144
3.2.2.	Estimación de los coeficientes en la descomposición K-L . . . . .	147
3.2.3.	Estimación de la densidad espectral de potencia . . . . .	150
3.2.4.	Estimación de parámetros de regresión . . . . .	153
3.3.	Estimación espectral no paramétrica . . . . .	157
3.3.1.	Método de los periodogramas . . . . .	157
3.3.2.	Algoritmo de cálculo del método de periodograma . . . . .	164
3.3.3.	Ventanas de estimación espectral . . . . .	166
3.4.	Estimación espectral paramétrica . . . . .	169
3.4.1.	Reacción de un sistema lineal a una señal estacionaria . . . . .	169
3.4.2.	Modelos paramétricos de señales aleatorias . . . . .	170
3.4.3.	Estimación de los parámetros a partir de la función de correlación . . . . .	172
<b>4.</b>	<b>Procesos de Markov</b>	<b>177</b>
4.1.	Definición y clasificación . . . . .	177
4.1.1.	Cadenas de Markov . . . . .	178
4.1.2.	Procesos discretos de Markov . . . . .	183
4.1.3.	Procesos continuos de Markov . . . . .	187
4.1.4.	Sucesiones de Markov . . . . .	198
4.2.	Análisis experimental de procesos de Markov . . . . .	201
4.2.1.	Modelos ocultos de Markov . . . . .	201
4.2.2.	Entrenamiento del modelo . . . . .	203
4.2.3.	Clasificación y comparación modelos . . . . .	212
<b>5.</b>	<b>Transformación de aleatoriedad</b>	<b>225</b>
5.1.	Paso de señales aleatorias por sistemas lineales . . . . .	225
5.1.1.	Análisis en el tiempo . . . . .	225
5.1.2.	Análisis en la frecuencia . . . . .	232
5.1.3.	Empleo de operadores lineales . . . . .	234
5.2.	Paso de señales aleatorias por sistemas no lineales . . . . .	237
5.2.1.	Métodos de linealización . . . . .	237
5.2.2.	Series de Volterra . . . . .	243
5.3.	Transformación de procesos de Markov . . . . .	249
5.3.1.	Transformación de señales aleatorias reales por procesos de Markov . . . . .	249
5.3.2.	Modelos de transformación de señales . . . . .	252
5.3.3.	Ejemplos de procesos generados de Markov . . . . .	254
5.3.4.	Discretización de procesos continuos de Markov . . . . .	257
<b>6.</b>	<b>Detección y filtración</b>	<b>265</b>
6.1.	Métodos de detección . . . . .	266
6.1.1.	Detección bayesiana de señales . . . . .	267
6.1.2.	Detección de máxima verosimilitud . . . . .	273
6.2.	Estimación de parámetros en señales aleatorias . . . . .	275
6.2.1.	Estimación en condiciones de ruido aditivo . . . . .	275
6.2.2.	Estimación de máxima verosimilitud . . . . .	275
6.2.3.	Combinación lineal de señales . . . . .	277
6.3.	Filtración óptima lineal por mínimos cuadrados . . . . .	281

---

6.3.1. Optimización de la respuesta a impulso . . . . .	281
6.3.2. Condición de realización física . . . . .	285
6.3.3. Filtros acoplados . . . . .	289
<b>7. Sistemas estocásticos . . . . .</b>	<b>295</b>
7.1. Modelos de sistemas dinámicos con entradas aleatorias . . . . .	295
7.1.1. Sistemas lineales estacionarios . . . . .	295
7.1.2. Variables de estado y modelado de procesos aleatorios . . . . .	296
7.1.3. Estimación de parámetros en sistemas estocásticos . . . . .	301
7.1.4. Identificación de parámetros en sistemas lineales . . . . .	303
7.2. Filtración de sistemas dinámicos . . . . .	307
7.2.1. Filtración Wiener . . . . .	307
7.2.2. Filtración óptima de sistemas discretos lineales . . . . .	308
7.2.3. Filtración óptima de sistemas continuos lineales . . . . .	312
7.2.4. Filtración de sistemas no lineales . . . . .	316
7.2.5. Filtración adaptativa de sistemas dinámicos . . . . .	318
<b>Bibliografía . . . . .</b>	<b>323</b>
<b>Índice alfabético . . . . .</b>	<b>327</b>

# Notaciones

Notación	Significado
$x, \mathbf{x}$	Escalar, Vector escalar
$x(s), x^*(s)$	Función con variable $s$ , Conjugado de $x$ por $s$
$x(s_k)$	Valor de la función en el argumento $s_k$
$f_{a,b}(s), \mathbf{f}$	Función $f$ con parámetros $a, b$ y argumento $s$ , vector funcional
$\{x_k : k = 1, \dots, N\}$	Serie o sucesión de valores con volumen $N$
$x[k]$	Señal discretizada (base normalizada $k$ )
$\langle x, y \rangle$	Producto interno de $x$ e $y$
$d(x_m, y_n), d(x, y)$	Distancia entre los elementos $x_m$ e $y_n$ , o entre las funciones $x$ e $y$
$\ x\ $	Norma de $x$
$x_k(t)$	Trayectoria $k$ de la señal $x(t)$
$\dot{x}(s), \ddot{x}(s)$	Primera, segunda derivada de la función $x$ por $s$
$\text{rank}(\mathbf{x}), \text{rank}(\mathbf{X})$	Rango del vector $\mathbf{x}$ , de la matriz $\mathbf{X}$
$\mathcal{E}_f, \mathcal{P}_f$	Energía de la señal $f$ , Potencia de la señal $f$
$\mathbf{Z}_{m \times n}, \mathbf{\Phi}^\top$	Matriz de orden $m \times n$ , Matriz transpuesta
$\mathbf{J}(x_1, \dots, x_n)$	Jacobiano por las variables $x_1, \dots, x_n$
$\text{supp}(x), \text{car}(x), \text{trace}(\mathbf{X})$	Soporte de la función $x$ , cardinal de $x$ , traza de $\mathbf{X}$
$\mathbf{I}, \mathbf{i}$	Matriz unitaria, vector unitario
$\mathfrak{X}$	Conjunto, espacio
$\mathcal{O}\{N\}$	Orden del número de operaciones
$o(x)$	Función de valores despreciables, $o(x)/x \rightarrow 0, x \rightarrow 0$
$\mathcal{K}\{x\}(s)$	Transformada (operador) sobre $x$ con argumento $s$
$\mathcal{F}, \mathcal{Z}, \mathcal{L}$	T. Fourier, Zeta, Laplace
$X(s)$	Representación espectral con variable $s$
$\mathbb{Z}, \mathbb{N}, \mathbb{R}, \mathbb{C}$	Dominio de los enteros, naturales, reales y complejos
$\Re\{x\}, \Im\{x\}$	Parte real de $x$ , parte imaginaria de $x$
$\xi(s)$	Señal aleatoria (alfabeto griego) de variable $s$
$\mathcal{B}(\mathfrak{X})$	$\sigma$ -álgebra $\mathcal{B}$ de subconjuntos de Borell sobre el espacio $\mathfrak{X}$
$F_\xi(x), P(x)$	Función de distribución integral, Función de probabilidad
$p_\xi(x), p_\xi(x; y)$	FDP de la variable aleatoria $\xi$ , FDP con argumentos $x$ e $y$
$\overline{E\{\xi^n\}}$	Valor esperado de orden $n$ (promedio de ensamble)
$\xi^n(t)$	Valor esperado de orden $n$ (promedio de tiempo)
$m_{n\xi}, \mu_{n\xi}$	Momento inicial, centralizado de orden $n$ de la variable $\xi$
$\sigma_\xi^2$	Varianza de la variable $\xi$
$\Theta_\xi(\omega)$	Función característica de la variable $\xi$
$\tilde{\xi}$	Valor estimado de $\xi$
$R_{\xi,\eta}(\cdot), K_{\xi,\eta}(\cdot)$	Función de correlación y covarianza mutua entre $\xi$ y $\eta$
$S_\xi(\omega)$	Espectro de potencia de $\xi$
$\Lambda_\xi$	Relación de verosimilitud para la variable aleatoria $\xi$
$\mathcal{N}, \mathcal{U}(m_{1\xi}, \sigma_\xi)$	Distribución (Gaussiana)(uniforme) con media $m_{1\xi}$ y varianza $\sigma_\xi^2$

## Acrónimos

TF	Transformada de Fourier
TIF	Transformada Inversa de Fourier
DEE	Densidad espectral de energía
DEP	Densidad espectral de potencia
TFTC	Transformada de Fourier en tiempo corto
TL	Transformada de Laplace
TW	Transformada wavelet
FDP	Función densidad de probabilidad
K-L	Karhunen-Loève (Transformada)
RBG	Ruido blanco Gaussiano



# Capítulo 1

## Representación de señales y sistemas

La representación, entendida como una idea que sustituye a la realidad, en el caso de señales y sistemas se orienta a la descripción de sus principales propiedades de interacción. En este capítulo se analizan los métodos de representación discreta e integral como formas fundamentales de la descripción de señales y sistemas.

### 1.1. Representación discreta de señales

**Definición 1.1.** Un espacio métrico corresponde a la pareja  $\mathfrak{X} = (X, d)$ , que consta del conjunto  $X$  o espacio de elementos (puntos) y la distancia  $d$ , que es una función real positiva semidefinida, que para cualquier pareja de elementos  $x, y \in X$ , cumple los axiomas:

1.  $d(x, y) = 0$ , si y solo si,  $x = y$ ; elementos con propiedades geométricas iguales.
2.  $d(x, y) = d(y, x)$ ; axioma de simetría
3.  $d(x, z) \leq d(x, y) + d(y, z)$ ; axioma triangular

#### 1.1.1. Espacio de representación de señales

En general, un conjunto de señales de análisis es el constituido por todas las funciones de valor complejo definidas en forma continua sobre un eje real, por ejemplo el del tiempo,  $L = \{x = x(t) : x(t) \in \mathbb{C}, t \in \mathbb{R}\}$ , donde  $L = \{x : \mathcal{K}\}$  es el conjunto formado por todos los  $x$ , para los cuales el atributo o relación  $\mathcal{K}$  es cierto:  $\mathcal{K} \Rightarrow x \in L$ . La mayoría de los espacios de funciones de señales se restringen a los espacios convencionales de Lebesgue:

$$L^p = L^p(\mathbb{R}) = \left\{ x \in L : \|x\|_p = \left( \int_{\mathbb{R}} |x(t)|^p dt \right)^{1/p} < \infty \right\}, p \geq 1 \quad (1.1)$$
$$L^\infty = L^\infty(\mathbb{R}) = \left\{ x \in L : \|x\|_\infty = \sup_t |x(t)| < \infty \right\}$$

donde  $\|\mathbf{x}\|_p$  es la *norma* definida para  $\mathbf{x}$  en el espacio  $L^p$ . Un espacio lineal  $L$ , en el cual se define una norma, se denomina *normalizado*. La restricción de  $p \geq 1$ ,  $p \in \mathbb{Z}$ , en (1.1), implica que la clase  $L^p(\mathbb{R})$  es un espacio normalizado, el cual es *completo* con respecto a la respectiva norma, esto es, cualquier sucesión fundamental de elementos converge.

La generación de espacios de señales, a partir de alguna condición común con interpretación física (energía, transformación a algún espacio complejo, etc.), implica establecer el modelo matemático formal de relación  $\mathcal{X}$  entre los elementos del conjunto. Un conjunto de funciones provisto de una distancia adecuada, conforma un *espacio de señales*. Un ejemplo de distancia entre dos señales,  $\mathbf{x}(t)$ ,  $\mathbf{y}(t) \in \mathbb{C}$ , definidas en  $T$ , es:

$$d(\mathbf{x}, \mathbf{y}) = \|\mathbf{x} - \mathbf{y}\|_2 = \left( \int_T |\mathbf{x}(t) - \mathbf{y}(t)|^2 dt \right)^{1/2} \quad (1.2)$$

El conjunto de funciones relacionados por la distancia (1.2), denominado espacio con *métrica cuadrática*, para los cuales la respectiva norma sea acotada,  $\|\mathbf{x}\|_2 < \infty$ , o *espacio*  $L^2(T)$  que es un espacio de Hilbert, está provisto del siguiente producto interno:

$$\langle \mathbf{x}, \mathbf{y} \rangle_{L^2(T)} = \int_T \mathbf{x}^*(t) \mathbf{y}(t) dt$$

El espacio de funciones, dado por la distancia (1.2), tiene amplio uso en la representación de señales debido a la interpretación física simple de su norma, la cual se asocia a la energía de las señales. Cuando  $\mathbf{x} \in L^2(T)$ , se tiene que,  $\|\mathbf{x}\|_2 = \langle \mathbf{x}, \mathbf{x}^* \rangle_{L^2(T)} \triangleq \mathcal{E}_x$ , que es la *energía* de la señal. Por cierto, la señal  $\mathbf{x}(t)$  se denomina *señal de energía* cuando

$$\mathcal{E}_x = \|\mathbf{x}\|_2 = \int_T |\mathbf{x}(t)|^2 dt < \infty, \quad (1.3)$$

La representación discreta de una señal  $\mathbf{x} \in L^2(T)$  de energía, que cumpla (1.3), implica hallar la transformación del espacio  $L^2(T)$  en el espacio  $\mathbb{C}^n$ , donde el valor de  $n$  se elige a partir del compromiso entre la precisión y la economía de la representación. La forma general para hallar esta representación consiste en la selección de un subespacio con dimensión  $n$  a partir de  $L^2(T)$ . Teniendo en cuenta que  $L^2(T)$  es un espacio completo separable [1], la señal  $\mathbf{x} \in L^2(T)$  puede ser representada de manera aproximada con cualquier precisión, si la dimensión de representación se escoge suficientemente grande, por medio de un conjunto de valores o coeficientes  $\mathbf{x}_k$ , ( $k \rightarrow \infty$ ), expresados en combinación lineal del siguiente espacio de funciones ordenadas, que se elige adecuadamente:

$$\mathbf{x}(t) = \sum_k^{\infty} \mathbf{x}_k \phi_k(t), \quad (1.4)$$

donde  $\phi_k(t)$  es el conjunto de funciones elegidas a priori, que conforman una base del espacio vectorial  $L^2(T)$ , las cuales son denominadas *funciones base*.



### 1.1.2. Descomposición en funciones ortogonales

**Definición 1.2.** Un espacio de señales complejas  $\{\phi_k(t) \in \mathbb{C}\}$  se define *ortogonal*, dentro del intervalo de representación  $(t_i, t_f)$ , si para el producto interno  $\langle \phi_m, \phi_n^* \rangle$  se cumple:

$$\langle \phi_m, \phi_n^* \rangle = \int_{t_i}^{t_f} \phi_m(t) \phi_n^*(t) dt = \int_{t_i}^{t_f} \phi_m^*(t) \phi_n(t) dt = \begin{cases} 0, & m \neq n \\ \mathcal{E}_m, & m = n \end{cases} \quad (1.5)$$

siendo  $\mathcal{E}_m = \mathcal{E}_n = \|\phi\|_2$  un valor de energía.

Como se deduce de la expansión ortogonal (1.4), una señal de energía  $x(t)$  se puede representar de forma aproximada en términos de un conjunto de funciones base  $\{\phi_n(t) : n = 0 \dots, N\}$ , para el cual los coeficientes  $x_n$  caracterizan el peso de la correspondiente función ortogonal  $\phi_n(t)$ . Los valores  $x_n$  se determinan de acuerdo con la condición del mínimo error cuadrático medio:

$$\overline{\varepsilon_N^2(t)} = \frac{1}{t_f - t_i} \int_{t_i}^{t_f} \left| x(t) - \sum_{n=0}^N x_n \phi_n(t) \right|^2 dt, \text{ siendo } \overline{\varepsilon_N^2(t)} \geq 0 \quad (1.6)$$

Como resultado se obtienen los siguientes coeficientes de descomposición:

$$x_n = \frac{\langle x, \phi_n \rangle}{\|\phi\|_2} = \frac{\int_{t_i}^{t_f} x(t) \phi_n^*(t) dt}{\int_{t_i}^{t_f} |\phi_n(t)|^2 dt} = \frac{1}{t_f - t_i} \int_{t_i}^{t_f} x(t) \phi_n^*(t) dt \quad (1.7)$$

Una señal de potencia (o energía) se descompone mediante un conjunto de funciones base, que genera un espacio completo, cuando los coeficientes  $x_n$  se determinan por (1.7).

El conjunto de los coeficientes  $\{x_n\}$  se denomina *espectro de la señal  $x(t)$* , mientras cada producto  $x_n \phi_n(t)$  se define como su respectiva componente espectral. Cualquiera de las dos formas, la serie generalizada de Fourier (1.4) o *el espectro*  $\{x_n\}$  (1.7), determinan unívocamente la señal  $x(t)$ .

### 1.1.3. Ejemplos de conjuntos ortogonales completos

El conjunto ortogonal más conocido corresponde a la *serie exponencial de Fourier*, que se define por las funciones exponenciales complejas del tipo:

$$\phi_n(t) = e^{jn\omega_0 t}, \quad (1.8)$$

donde  $\omega_0 = 2\pi/\mathcal{E}_n$ , siendo  $\mathcal{E}_n = (t_f - t_i)$  y  $n \in \{0, \pm 1, \pm 2, \dots\}$ , que corresponde a cada término de la serie o *armónico*.

Cabe anotar, que todas las expresiones anteriores también son válidas para la representación espectral de las funciones periódicas, reemplazando el intervalo de definición de las señales aperiódicas  $[t_i, t_f]$  por el valor del período  $T$  de las funciones periódicas.

En la Tabla (1.1) se muestran los coeficientes de Fourier para algunas señales comunes.

Señal	Definición	Coeficiente $x_n$
Cuadrada simétrica	$\begin{cases} 1, &  t  < \frac{T}{4} \\ -1, & \frac{T}{4} <  t  < \frac{T}{2} \end{cases}$	$\begin{cases} \text{sinc}(n\pi/2), & n \neq 0 \\ 0, & n = 0 \end{cases}$
Pulsos rectangulares	$\begin{cases} +1, &  t  < \frac{\tau}{2} \\ 0, & \frac{\tau}{2} \leq  t  < \frac{T}{2} \end{cases}$	$\frac{\tau}{T} \text{sinc}(n\pi\tau/T)$
Triangular simétrica	$1 - 4 t /T, \quad  t  < T/2$	$\begin{cases} \text{sinc}(n\pi/2), & n \neq 0 \\ 0, & n = 0 \end{cases}$

**Tabla 1.1.** Coeficientes de descomposición espectral de Fourier

Se considera que los motivos principales, por los cuales las series de Fourier tienen amplia aplicación, son los siguientes [2]:

- Se representan mediante oscilaciones simples determinadas para todo valor de  $t$ .
- Se descomponen en oscilaciones armónicas **sin** y **cos**, que son las únicas funciones del tiempo que conservan su forma al pasar por cualquier dispositivo lineal; sólo varían su fase y su amplitud.
- La descomposición en senos y cosenos permite aplicar el método simbólico usualmente desarrollado para el análisis de circuitos lineales.

En la aproximación de señales continuas se emplean diferentes polinomios y funciones especiales ortogonales. Escogiendo adecuadamente estas funciones en la descomposición ortogonal (1.4), se obtiene la representación aproximada para la señal a partir de un número pequeño de elementos de la serie.

De otra parte existen funciones, que conforman sistemas base completos  $\{\phi_n(t)\}$ , siendo ortonormales en los intervalos indicados con valor de peso  $\sqrt{w(t)}$ , esto es, que se cumpla la condición:

$$\{\phi_n(t)\} = \left\{ \sqrt{w(t)}\psi_n(t) \right\}, \quad \forall n \quad (1.9)$$

donde  $w(t)$  es la *función de peso* y  $\{\psi_n(t)\}$  son las funciones sobre los cuales se forman los sistemas base ortogonales. Como consecuencia de (1.9), la definición de ortogonalidad (1.5) sobre el intervalo  $T$ , cambia por:

$$\int_T \psi_m(t) \psi_n^*(t) w(t) dt = \begin{cases} 0, & m \neq n \\ \mathcal{E}_n, & m = n \end{cases}, \quad \mathcal{E}_n = \left\| \sqrt{w(t)}\psi_n(t) \right\|_2 = \int_T \left( \sqrt{w(t)}\psi_n(t) \right)^2 dt$$

Así mismo, cambia la definición de los coeficientes (1.7) de la expansión (1.4):

$$x_n = \frac{1}{\left\| \sqrt{w(t)} \psi_n(t) \right\|_2} \int_T x(t) \sqrt{w(t)} \psi_n^*(t) dt. \quad (1.10)$$

Particular importancia tienen las funciones de Walsh, para las cuales  $\kappa(l)$  corresponde al elemento  $l$  en representación del valor  $\kappa$  en código Gray, mientras  $r_n(t)$  son las funciones de Rademascher [3]. La descomposición de una señal,  $x(t)$ , en este caso tiene la forma:

$$x(t) = \sum_{n=0}^{\infty} x_n \text{wal}_n(t) \quad (1.11)$$

donde los coeficientes de la serie Fourier-Walsh (1.11), que se determinan de acuerdo con (1.7) y (1.10),

$$x_n = \int_{-1/2}^{1/2} x(t) \text{wal}_n(t) dt$$

conforman el espectro de la señal  $x(t)$  por la base Walsh, el cual es denominado *S-espectro* (*sequency spectrum*).

**Ejemplo 1.1.** Determinar el *S*-espectro de la señal  $x(t) = \sin 2\pi t$ ,  $-\infty < t < \infty$ , empleando el conjunto  $\{\text{wal}_n(t), 0 \leq t < 1, n = 0, 1, \dots, 31\}$ .

Debido a la imparidad de la función  $x(t)$  respecto al punto  $t = 1/2$ , todos los coeficientes para las funciones pares  $\text{wal}_{2n}(t) = 0$ , al igual que los coeficientes de las funciones impares 3, 7, 11, 15, 19, 23, 27, 31. Los demás coeficientes tienen los siguientes valores:

$$\begin{aligned} x_1 &= \int_{-1/2}^{1/2} x(t) \text{wal}_1(t) dt = \int_{-1/2}^{1/2} \sin(2\pi t) \text{wal}_1(t) dt = 2 \int_0^{1/2} \sin(2\pi t) dt = \frac{2}{\pi} \approx 0.637 \\ x_5 &= \int_{-1/2}^{1/2} \sin(2\pi t) \text{wal}_5(t) dt = 4 \int_0^{1/8} \sin(2\pi t) dt - 2 \int_{1/8}^{3/8} \sin(2\pi t) dt \approx -0.264 \end{aligned}$$

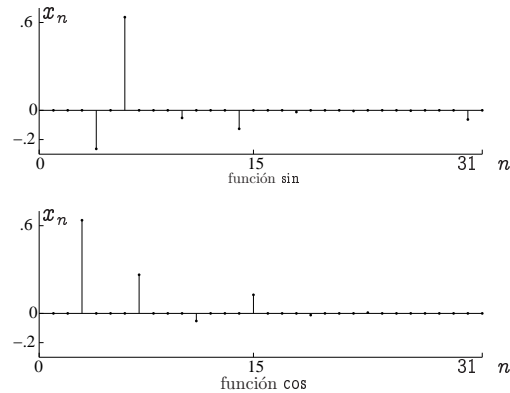
En forma similar, se obtienen los siguientes coeficientes:

$$\begin{aligned} x_9 &\approx -0.052, \quad x_{13} \approx -0.127, \quad x_{17} \approx -0.012, \quad x_{21} \approx -0.005, \\ x_{25} &\approx -0.026, \quad x_{29} \approx -0.063 \end{aligned}$$

De igual manera, se puede establecer el *S*-espectro de la señal  $x(t) = \cos 2\pi t$ , para la cual se obtiene los siguientes coeficientes:  $x_2 \approx 0.637$ ,  $x_6 \approx 0.264$ ,  $x_{10} \approx -0.052$ ,  $x_{14} \approx 0.127$ ,  $x_{18} \approx -0.012$ ,  $x_{22} \approx 0.005$ . Los demás coeficientes son iguales a cero. En la Figura 1.1 se observan las respectivas representaciones.

En general, el *S*-espectro de la función  $x(t) = \sin(2\pi t + \theta)$  se obtiene a partir de la relación:

$$\sin(2\pi t + \theta) = \sin(2\pi t) \cos \theta + \cos(2\pi t) \sin \theta$$



**Figura 1.1.** Representación espectral por la base Walsh

por lo cual es claro, que cambiando el desfase  $\theta$  cambian los valores de los coeficientes de descomposición de la serie de Walsh.

## Problemas

**Problema 1.1.** Sean las señales  $y(t)$  y  $x(t)$  elementos de un espacio real de Hilbert, tales que son linealmente independientes, y por tanto la igualdad  $y = \lambda x$  no tiene lugar para ningún valor real de  $\lambda$ . Demostrar la desigualdad de Cauchy-Buniakovski,  $|\langle x, y \rangle| < \|x\| \cdot \|y\|$ .

**Problema 1.2.** En un espacio de Hilbert están dados los vectores  $u$  y  $v$ , tal que  $\|v\| = 1$ . En analogía con la geometría de los vectores comunes en un plano, el vector  $w = \langle u, v \rangle$  se denomina *proyección ortogonal* del vector  $u$  en la dirección  $v$ . Demostrar que el vector  $y = u - w$  es ortogonal al vector  $v$ .

**Problema 1.3.** Demostrar la ortogonalidad del conjunto formado por el principio  $\text{sgn}[\sin(2^m \pi \theta)]$ ,  $m \in \mathbb{Z}$ .

**Problema 1.4.** Una señal compleja periódica  $x(t)$  en el intervalo de tiempo  $-T/2 \leq t \leq T/2$  tiene la forma,  $x(t) = x_1(t) + jx_2(t)$ . Demostrar que si la función  $x_1(t)$  es par, mientras la función  $x_2(t)$  es impar, entonces los coeficientes  $x_n$  de la serie de Fourier, para cualquier  $n$ , serán valores reales.

**Problema 1.5.** Calcular los coeficientes de la serie compleja de Fourier para las siguientes funciones periódicas dadas en el intervalo  $[-\pi, \pi]$ :

a).  $x(t) = t$ . b).  $x(t) = |\sin t|$ . c).  $x(t) = \text{sgn}(t)$ .

**Problema 1.6.** Determinar el espectro de la serie Haar-Fourier para la sucesión periódica de pulsos cuadrados con amplitud  $a = 2^m$ ,  $m \in \mathbb{Z}$ , apertura  $\tau = 1/a$ , y periodo unitario, empleando el sistema  $\{\text{har}_n(t), 0 \leq t < 1, n = 0, 1, 2, \dots, N, N = 2^k - 1, k \in \mathbb{Z}\}$  siendo  $m = 3$  y  $k = 4$ .

**Ejercicio en el CP 1.1.** Desarrollar un programa gráfico que convierta la serie trigonométrica de Fourier utilizando desde 1 hasta 15 armónicos y 50 puntos por periodo.

## 1.2. Representación integral de señales y sistemas

### 1.2.1. Densidad espectral

La representación discreta, analizada en el numeral §1.1, sirve en el caso de descomposición de señales periódicas o de las señales aperiódicas, definidas sobre intervalos finitos de tiempo. Esta última restricción es superable al emplear la representación integral, la cual resulta en forma de transformadas integrales, que se lleva a cabo para conjuntos base continuos. Específicamente, la generalización de (1.4), para cualquier conjunto base o *transformada directa generalizada de Fourier*, se puede escribir en el dominio  $T$  como sigue:

$$v(s) = \int_T x(t) \theta(s, t) dt, \quad s \in S \quad (1.12)$$

siendo  $s$  el parámetro generalizado del dominio  $S$  y  $\theta(s, t)$  el *núcleo base conjugado*. La función  $v(s)$ , que es una representación continua de  $x(t)$  similar a los coeficientes  $x_n$  en (1.4), corresponde a la *densidad*, la cual caracteriza la distribución de  $x(t)$  con relación a  $\theta(s, t)$  en las diferentes regiones de  $S$ . De forma similar, se determina la *transformada inversa generalizada de Fourier*:

$$x(t) = \int_S v(s) \phi(s, t) ds, \quad t \in T \quad (1.13)$$

Al reemplazar  $x(t)$  de (1.13) en (1.12) e intercambiando el orden de integración se obtiene la condición conjunta que cumplen los núcleos base conjugados:

$$x(t) = \int_S \int_T x(\tau) \theta(s, \tau) \phi(t, s) d\tau ds = \int_T k(t, \tau) x(\tau) d\tau = f_k \{x(t)\} \quad (1.14)$$

donde el núcleo de la integral se define como:

$$k(t, \tau) = \int_S \theta(s, \tau) \phi(t, s) ds$$

En forma general, el funcional lineal  $f_k \{x(t)\}$  no es acotado ni continuo, por lo tanto no es de esperar que exista una función, perteneciente a  $L^2(T)$  y definida en  $\tau$ , que cumpla la condición (1.14) para cualquier señal  $x(t)$ . La solución a este problema está en la representación de  $x(t)$  mediante la función generalizada singular  $\delta(t)$  de la forma [4]:

$$f_k \{x(t)\} = x(t) = \int_T \delta(t - \tau) x(\tau) d\tau, \quad t \in T$$

Dado un núcleo base, de (1.14) se establece la restricción para su núcleo conjugado:

$$\int_S \phi(t, s) \theta(s, \tau) ds = \delta(t - \tau) \quad (1.15)$$

De la misma manera, al reemplazar (1.12) en (1.13) se tiene la condición:

$$\int_T \theta(s, t) \phi(t, \sigma) dt = \delta(s - \sigma) \quad (1.16)$$

que debe cumplir tanto  $\theta(s, t)$ , como  $\phi(t, \sigma)$  para conformar un par de núcleos conjugados base. Sin embargo, el análisis de (1.15) y (1.16) muestra que ambos núcleos al mismo tiempo no pueden pertenecer a  $L^2(\mathbb{R})$ . Algunos núcleos base tienen la propiedad por la cual la transformada de la función en el dominio  $t \in T$ , siendo integrable y perteneciente a  $L^2(T)$ , siempre genera funciones en el dominio  $s \in S$  también integrables (pertenecientes a  $L^2(S)$ ). La correspondencia completa entre las funciones determinadas en los dominios de  $t$  y  $s$  se provee mediante núcleos autoconjugados (también denominados *inversos o mutuos*), esto es [4],  $\theta^*(s, t) = \phi(t, s)$ .

La transformada integral de Fourier, o *Transformada de Fourier* (TF), se obtiene asumiendo en calidad de base (1.8) la función,  $\phi(t, s) = e^{j2\pi st}$  a la cual le corresponde el núcleo base conjugado,  $\theta(t, s) = e^{-j2\pi st}$ , definido para los argumentos  $T \subset (-\infty, \infty)$  y  $S \subset (-\infty, \infty)$ , que corresponde a la *Transformada Inversa de Fourier* (TIF):

$$x(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} X(\omega) e^{j\omega t} d\omega \triangleq \mathcal{F}^{-1}\{X(\omega)\} \quad (1.17)$$

De igual manera, se puede obtener la *Transformada Directa de Fourier* (TDF):

$$X(\omega) = \int_{-\infty}^{\infty} x(t) e^{-j\omega t} dt \triangleq \mathcal{F}\{x(t)\} \quad (1.18)$$

El par de transformadas (1.17) y (1.18) representa la señal  $x(t)$  como una suma continua de funciones exponenciales cuyas frecuencias están en el intervalo  $(-\infty, \infty)$ .

La transformada de Fourier  $X(\omega)$  corresponde a la función de *densidad espectral de energía* (DEE), cuya existencia depende de las *condiciones de Dirichlet* [5]:

1.  $x(t)$  tiene un número finito de máximos y mínimos en cualquier intervalo de tiempo finito,
2.  $x(t)$  tiene un número finito de discontinuidades finitas en cualquier intervalo de tiempo finito,
3.  $x(t)$  es absolutamente integrable.

Con relación a la última condición, se puede decir que la transformada de Fourier representa unívocamente cualquier señal real de energía. Así mismo, se considera que las condiciones de Dirichlet son suficientes para la convergencia de la TF en la representación de una señal de potencia en el intervalo  $(0, T)$ .

Las principales propiedades de la TF se muestran en la Tabla 1.2.

Propiedad	Descripción
Linealidad	$\mathcal{F}\left\{\sum_n a_n x_n(t)\right\} = \sum_n a_n X_n(\omega), \forall a_n = \text{const.}$
Conjugada compleja	$\mathcal{F}\{x^*(t)\} = X^*(-\omega)$
Dualidad	$\mathcal{F}\{X(t)\} = 2\pi x(-\omega)$
Escala de coordenadas	$\mathcal{F}\{x(\alpha t)\} = \frac{X(\omega/\alpha)}{ \alpha }$
Desplazamiento en el tiempo (retardo)	$\mathcal{F}\{x(t - t_0)\} = X(\omega) e^{-j\omega t_0}$
Desplazamiento de frecuencia (modulación)	$\mathcal{F}\{x(t) e^{\pm j\omega_0 t}\} = X(\omega \mp \omega_0)$
Diferenciación	$\mathcal{F}\left\{\frac{d}{dt}x(t)\right\} = j\omega X(\omega)$
Integración	$\mathcal{F}\left\{\int_{-\infty}^t x(\tau) d\tau\right\} = \frac{1}{j\omega}X(\omega) + \pi X(0)\delta(\omega)$
Teorema de Rayleigh	$\mathcal{E}_x = \int_{-\infty}^{\infty} X(\omega) X(-\omega) d\omega = \frac{1}{2\pi} \int_{-\infty}^{\infty}  X(\omega) ^2 d\omega$
Teorema de Parsevall	$\mathcal{E}_x = \int_{-\infty}^{\infty} X(\omega) X^*(\omega) d\omega$ $= \frac{1}{2\pi} \int_{-\infty}^{\infty}  X(\omega) ^2 d\omega = 2\pi \int_{-\infty}^{\infty} x^2(t) dt$
Multiplicación	$\mathcal{F}\{x_1(t) x_2(t)\} = X_1(\omega) * X_2(\omega)$
Convolución	$\mathcal{F}\{x_1(t) * x_2(t)\} = X_1(\omega) X_2(\omega)$

Tabla 1.2. Propiedades de la transformada de Fourier

**Ejemplo 1.2.** Determinar la DEE de la función pulso cuadrado  $x(t) = a \text{rect}_\tau(t)$ .

Al realizar el cálculo de la TDF, (1.18), se obtiene:

$$\begin{aligned}
 X(\omega) &= \int_{-\infty}^{\infty} a \text{rect}_\tau(t) e^{-j\omega t} dt = a \int_{-\tau/2}^{\tau/2} e^{-j\omega t} dt = \frac{a}{j\omega} (e^{j\omega\tau/2} - e^{-j\omega\tau/2}) = \frac{a\tau \sin(\omega\tau/2)}{\omega\tau/2} \\
 &= a\tau \text{sinc}(\omega\tau/2).
 \end{aligned}$$

En general, se pueden formar nuevos conjuntos base a partir del desplazamiento en el tiempo de un sistema base original, así

$$\phi(s, t) = \phi(t - s) \quad (1.19)$$

entonces, el sistema base es función de una sola variable, dada por la diferencia de ambos argumentos,  $(t - s)$ .

Si se considera  $T, S \in (-\infty, \infty)$ , entonces, la señal original es la *integral de convolución*:

$$x(t) = \int_{-\infty}^{\infty} v(s) \phi(s-t) ds \triangleq v(t) * \phi(t) \quad (1.20)$$

El núcleo conjugado, en caso de existir, se puede hallar mediante la TF de (1.20), que aplicando la propiedad de convolución en el tiempo, se obtiene que  $X(\omega) = V(\omega) \Phi(\omega)$ , de lo que resulta que  $V(\omega) = \Theta(\omega) X(\omega)$ , donde  $\Theta(\omega) = 1/\Phi(\omega)$ . Por lo tanto, el núcleo conjugado también depende de la diferencia de argumentos

$$v_x(s) = \int_{-\infty}^{\infty} x(t) \theta(t-s) dt \quad (1.21)$$

En calidad de ejemplo concreto, en (1.19), se puede analizar el desplazamiento de la función base  $\phi(t-s) = \delta(t-s)$ , entonces  $\Phi(f) = \Theta(f) = 1$ , con lo cual se obtiene que

$$v_x(s) = \int_{-\infty}^{\infty} x(t) \delta(t-s) dt = x(s) \quad (1.22)$$

**Transformada de Laplace.** El análisis de algunas señales  $x(t)$ , mediante la TF se dificulta cuando no cumplen la condición débil de Dirichlet, esto es, no son absolutamente integrables. En este caso, se puede hallar la convergencia, si la señal se multiplica por una función que acote los valores extremos, por ejemplo, por una curva exponencial del tipo,  $e^{-ct}$ , en la cual la constante  $c > 0$  se escoge de tal manera que asegure la condición de integrabilidad absoluta del producto  $x(t)e^{-ct}$ , entonces, la DEE, (1.17), toma la forma,

$$X(\omega, c) = \int_{-\infty}^{\infty} (x(t) e^{-ct}) e^{-j\omega t} dt \quad (1.23)$$

Sin embargo, para asegurar la convergencia de (1.23), se debe ajustar la señal en el tiempo,  $x(t)$ , tal que  $\forall t < 0$ , su aporte sea 0. De otra manera, el multiplicando  $e^{-ct}$  puede conllevar a la divergencia de la integral. Por lo tanto, el límite inferior de la densidad acotada siempre es 0:

$$X(\omega, c) = \int_0^{\infty} (x(t) e^{-ct}) e^{-j\omega t} dt = \int_0^{\infty} x(t) e^{-(c+j\omega)t} dt = X(c+j\omega) \quad (1.24)$$

Si se halla la TF inversa de la densidad espectral  $X(c+j\omega)$ , se obtiene que

$$\frac{1}{2\pi} \int_{-\infty}^{\infty} X(c+j\omega) e^{j\omega t} d\omega = x(t) e^{-ct}$$



Así mismo, si ambas partes de la integral anterior se multiplican por  $e^{ct}$ , juntando los factores de multiplicación exponencial y haciendo el cambio de variable  $\omega \rightarrow c + j\omega$ , entonces, se llega a la expresión:

$$\frac{1}{2\pi j} \int_{c-j\infty}^{c+j\infty} X(c+j\omega) e^{(c+j\omega)t} d(c+j\omega) = x(t)$$

En la práctica, es usual el empleo de la notación  $p \triangleq c + j\omega$ , entonces la anterior integral y la densidad (1.24), conforman el par de *Transformadas de Laplace* (TL):

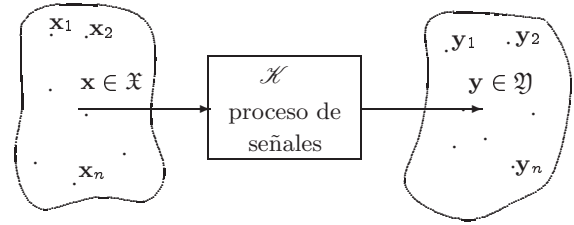
$$\int_0^{\infty} x(t) e^{-pt} dt = X(p) \triangleq \mathcal{L}\{x(t)\} \quad (1.25a)$$

$$\frac{1}{2\pi j} \int_{c-j\infty}^{c+j\infty} X(p) e^{pt} dp = x(t) \triangleq \mathcal{L}^{-1}\{X(p)\} \quad (1.25b)$$

De (1.25b), se observa que la TL inversa se realiza mediante la integración en el plano complejo  $p$  sobre toda la recta  $c = \text{const.}$  También es claro, que el reemplazo de  $p = j\omega$  conduce directamente a la TF, por lo que, todas sus propiedades tienen lugar para la TL.

### 1.2.2. Transformaciones lineales

A su paso por los sistemas, las señales sufren cambios en cada una de las etapas del proceso. Debido a la diversidad de formas posibles de señal es preferible tener una representación común para su análisis. La transformación de la señal se puede analizar como el enlace entre la entrada y la salida de un sistema o circuito (Figura 1.2), pues en sí corresponde a la representación de un conjunto inicial de señales  $\mathfrak{X}$  en términos de otro conjunto; el de salida  $\mathfrak{Y}$ .



**Figura 1.2.** Transformación de señales

En forma general, para la descripción de la transformación  $\mathcal{K}$  se deben conocer todas las posibles parejas ordenadas entrada - salida,  $\{\mathbf{x}, \mathcal{K}\{\mathbf{x}\}\}$ , lo cual en la mayoría de los casos prácticos no es realizable, entre otras razones, porque se puede tener una alta cantidad de enlaces, además, por la dificultad en su ordenamiento. Sin embargo, cuando se restringe el análisis a las transformadas lineales, la transformación se puede describir completamente empleando un subconjunto reducido de todas las posibles parejas entrada - salida, como consecuencia de la propiedad de superposición de las transformadas lineales. Los modelos lineales son el referente de sistemas más empleados, de hecho, es común el análisis de sistemas no lineales, a través de su linealización.

Una transformación lineal  $\mathcal{K}$ , que corresponde a la imagen determinada en un espacio lineal  $\mathfrak{X}$ , tiene las siguientes propiedades:

1. *Aditividad*,  $\mathcal{K}\{\mathbf{x}_1 + \mathbf{x}_2\} = \mathcal{K}\{\mathbf{x}_1\} + \mathcal{K}\{\mathbf{x}_2\}$ ,  $\forall \mathbf{x}_1, \mathbf{x}_2 \in \mathfrak{X}$ ,
2. *Homogeneidad*,  $\mathcal{K}\{\alpha \mathbf{x}\} = \alpha \mathcal{K}\{\mathbf{x}\}$ ,

o lo que es equivalente  $\mathcal{K}\{\alpha \mathbf{x}_1 + \beta \mathbf{x}_2\} = \alpha \mathcal{K}\{\mathbf{x}_1\} + \beta \mathcal{K}\{\mathbf{x}_2\}$ , donde  $\alpha, \beta = \text{const.}$  De lo anterior resulta que  $\mathcal{K}\{\mathbf{0}\} = 0$  y  $\mathcal{K}\{-\mathbf{x}\} = -\mathcal{K}\{\mathbf{x}\}$ , por lo tanto, el conjunto de vectores linealmente transformados corresponde a un espacio lineal con el mismo conjunto de escalas que el conjunto de definición de la transformada.

De otra parte, si los espacios de entrada y salida son idénticos, la transformación lineal se denomina *operador lineal*, esto es, cuando el dominio de representación original coincide con el dominio de la imagen de la transformada. En los operadores lineales la multiplicación de dos operadores se determina como:

$$\mathcal{K} = \mathcal{K}_1 \mathcal{K}_2 \Rightarrow \mathcal{K}\{\mathbf{x}\} = \mathcal{K}_1\{\mathcal{K}_2\{\mathbf{x}\}\}$$

que corresponde a la conexión en cascada de ambos operadores. Cabe anotar que el conjunto  $\mathbf{x} \in \mathfrak{X}$ , tal que  $\mathcal{K}\{\mathbf{x}\} = 0$ , se denomina núcleo o *kernel* ( $\ker \mathcal{K}$ ) del operador  $\mathcal{K}$ .

### Transformaciones lineales sobre espacios finitos

Un caso de análisis frecuente corresponde a las transformaciones lineales sobre subespacios funcionales de  $L^2(T)$  cuyos resultados se pueden generalizar a todo  $L^2(\mathbb{R})$ .

**Representación con vectores respuesta.** Sea el espacio de entrada  $\mathfrak{X}$  generado por la base linealmente independiente  $\{\varphi_i : i = 1, \dots, n\}$ , para la cual existe la base conjugada con elementos  $\{\theta_i : i = 1, 2, \dots, n\}$ . Sea el espacio imagen  $\mathfrak{Y}$  con dimensión  $n$  que contiene los valores de la transformación  $\mathcal{K}$ . Luego, a partir de la expansión generalizada de Fourier,  $\forall \mathbf{x} \in \mathfrak{X}$ , se obtiene que  $\mathbf{x}(t) = \sum_{i=1}^n \langle \mathbf{x}, \theta_i \rangle \varphi_i(t)$ ,  $t \in T$ , por lo que al considerar la linealidad de  $\mathcal{K}$ , entonces

$$\mathbf{y}(t) = \mathcal{K}\{\mathbf{x}(t)\} = \sum_{i=1}^n \langle \mathbf{x}, \theta_i \rangle \psi_i(t) \quad (1.26)$$

donde  $\{\psi_i : \psi_i(t) = \mathcal{K}\{\varphi_i(t)\}; i = 1, \dots, n\}$ , que son las respuestas para cada una de las funciones base del espacio  $\mathfrak{X}$ , dadas en calidad de entrada de  $\mathcal{K}$ . Luego, se puede considerar el conjunto  $\{\psi_i\}$  como la representación de  $\mathcal{K}$  en términos de la base  $\{\phi_i\}$  en  $\mathfrak{X}$ .

**Representación por sucesión de funcionales lineales.** Otra forma de representación de  $\mathcal{K}$  se da por la sucesión ordenada de  $n$  vectores en  $\mathfrak{X}$ . Sea  $\{\widehat{\varphi}_i : i = 1, \dots, n\}$  el conjunto de vectores linealmente independientes, que generan a  $\mathfrak{Y}$ , a los cuales le corresponde la base

conjugada  $\{\widehat{\theta}_i : i = 1, \dots, n\}$ . Entonces, cualquier vector de  $\mathfrak{Y}$  se representa como:

$$y(t) = \sum_{j=1}^n \langle \mathbf{y}, \widehat{\theta}_j \rangle \widehat{\varphi}_j(t) \quad (1.27)$$

teniendo en cuenta (1.26) y (1.27), se obtiene

$$y(t) = \mathcal{K} \{x(t)\} = \sum_{j=1}^n \langle \mathbf{x}, \phi_j \rangle \widehat{\varphi}_j(t), \quad \phi_i(t) = \sum_{j=1}^n \langle \widehat{\theta}_j, \psi_i \rangle \theta_i(t)$$

Por lo tanto, la sucesión ordenada  $\{\phi_j\}$  representa a  $\mathcal{K}$  en términos de la base dada de  $\mathfrak{Y}$ . Asumiendo la independencia lineal de  $\{\phi_j\}$ , entonces, la representación del operador se puede entender como una sucesión ordenada, en la cual cada elemento es una transformada de rango unitario (funcional lineal)  $f_j(\widehat{\mathbf{x}}) = \langle \mathbf{x}, \phi_j \rangle$ , con  $j = 1, \dots, n$ ; la suma por todos elementales representa toda la transformación:  $\mathcal{K} \{x(t)\} = \sum_{j=1}^n f_j(\mathbf{x}) \widehat{\varphi}_j(t)$ .

**Representación matricial.** Por cuanto los vectores  $\psi_i$  en el primer método, ó  $\phi_j$  en el segundo método, se describen por un conjunto de  $n$  coeficientes (vector fila) la transformación lineal puede ser representada por una tabla de relación de  $n \times n$  escalares, esto es, en forma de un arreglo matricial.

Sea

$$\psi_i = \sum_{j=1}^n \lambda_{ji} \widehat{\varphi}_j(t), \quad \lambda_{ji} = \langle \psi_i, \widehat{\theta}_j \rangle = \langle \mathcal{K} \{\varphi_i\}, \widehat{\theta}_j \rangle \quad (1.28)$$

Luego, al reemplazar (1.28) en (1.26) se tiene que:

$$y(t) = \sum_{i=1}^n \sum_{j=1}^n \lambda_{ji} \alpha_i \varphi_j(t) = \sum_{j=1}^n \beta_j \widehat{\varphi}_j(t), \quad (1.29)$$

donde  $\alpha_i = \langle \mathbf{x}, \theta_i \rangle$ , y  $\beta_i = \langle \mathbf{y}, \widehat{\theta}_i \rangle = \sum_{j=1}^n \lambda_{ji} \alpha_j$ .

La expresión (1.29) se escribe en forma matricial como  $\boldsymbol{\beta} = \mathbf{K}\boldsymbol{\alpha}$ , donde  $\boldsymbol{\alpha}$  y  $\boldsymbol{\beta}$  son vectores con dimensión  $n$  que representan a  $\mathbf{x}$  sobre  $\mathfrak{X}$  e  $\mathbf{y}$  sobre  $\mathfrak{Y}$  por sus respectivas bases. Cabe anotar que la transformación lineal  $\mathcal{K}$  se representa completamente mediante  $\mathbf{K}$ , que es una matriz de orden  $n$  con elementos

$$\lambda_{ji} = \langle \mathcal{K} \{\varphi_i\}, \widehat{\theta}_j \rangle \quad (1.30)$$

### 1.2.3. Representación integral de sistemas lineales

En el caso de transformaciones lineales, la descripción de la relación entrada-salida en grupos de reducidos elementos se puede generalizar a todo el sistema lineal empleando la propiedad de superposición. De otra parte, la presunción de linealidad en la transformación exige que la señal a la salida sea proporcional a la entrada, pero de ninguna manera que

dependa de procesos remanentes o transitorios existentes en el sistema. En lo sucesivo, el análisis de transformaciones se realiza para sistemas invariantes en el tiempo.

**Operadores lineales en espacios integrables.** El análisis de los sistemas puede llevarse a cabo mediante cualquiera de las distintas formas de representación de las señales. En el caso discreto, se puede emplear la representación de transformaciones lineales sobre espacios de dimensión finita, correspondientes al modelo de señal dado en (1.4). Sin embargo, se tiene que la implementación práctica de los métodos de análisis es compleja, debido a que se hace necesario una gran cantidad de mediciones para los espacios de entrada en la descripción de cada transformación; sumado al hecho de que la misma cantidad de transformaciones puede ser muy alta. Por tal razón, esta forma de análisis de sistemas es empleada en casos específicos de espacios de entrada con bajo número de dimensiones y transformaciones relativamente simples.

El análisis de sistemas es preferible llevarlo a cabo, basados en la representación integral de señales, empleando operadores lineales, los cuales se definen sobre la transformación lineal acotada y determinada en espacios integrables  $L^2(T)$ . La interpretación física de esta restricción tiene que ver con el cumplimiento de la condición de estabilidad de los sistemas, que implica que a una señal de entrada con energía le debe corresponder también otra señal de energía a la salida.

Empleando el modelo de bases continuas en la representación generalizada de Fourier (1.13), las respectivas señales de entrada y salida se describen como:

$$x(t) = \int_S v_x(s) \phi(t, s) ds, \quad (1.31a)$$

$$y(t) = \int_S v_y(s) \phi(t, s) ds. \quad (1.31b)$$

Así, se define el operador  $\mathcal{K}$ , tal que

$$y(t) = \mathcal{K} \{x(t)\} = \int_S v_y(s) \psi(t, s) ds$$

donde  $\psi(t, s) = \mathcal{K} \{\phi(t, s)\}$ .

La densidad de salida, que de acuerdo con (1.12) es  $v_y(s) = \int_T y(t) \theta(s, t) dt$  y al reemplazar (1.31b), resulta en

$$v_y(s) = \int_T \int_S v_x(\sigma) \psi(t, \sigma) \theta(s, t) d\sigma dt = \int_S K(s, \sigma) v_x(\sigma) d\sigma,$$

donde el núcleo de la anterior integral, definido como,

$$K(s, \sigma) = \int_T \psi(t, \sigma) \theta(s, t) dt \quad (1.32)$$

representa la transformación  $\mathcal{K}$  y corresponde al sentido de los coeficientes (1.30), en los cuales los coeficientes  $i$  y  $j$  se convierten en las variables  $\sigma$  y  $s$  de (1.32), mientras las expresiones de representación de entrada y salida, en términos de la base  $\varphi(t, s)$ , se obtienen de las respectivas representaciones integrales. Así por ejemplo, se puede tomar la base dada en (1.21), para la cual se cumple que  $\phi(t, s) = \delta(t - s)$ , con  $T, S \subset (-\infty, \infty)$ . En este caso, de acuerdo con (1.22), se tiene que  $v_x(s) = x(s)$  y  $v_y(s) = y(s)$ . La reacción a la función base, especialmente notada como

$$\psi(t, \sigma) \triangleq h(t, s) \quad (1.33)$$

se denomina *respuesta a impulso*, que como función en el tiempo describe la reacción del operador cuando a la entrada se tiene la función  $\delta$  dada en el momento de tiempo  $s$ . Luego, de (1.32) se tiene

$$K(s, \sigma) = \int_{-\infty}^{\infty} \delta(s - t) h(t, \sigma) dt = h(s, \sigma) \quad (1.34)$$

con lo cual, la salida (1.31b) se determina como

$$y(t) = \int_T h(t, \tau) x(\tau) d\tau \quad (1.35)$$

La expresión (1.35) es la primera forma de análisis de sistemas, que corresponde al empleo en calidad de núcleo de su reacción a la función  $\delta(t)$ , conocida como el *método de la integral de superposición*. En otras palabras, para describir el comportamiento de un sistema en el dominio del tiempo, se puede probar aplicando la función elemental  $\delta(t)$  a la entrada; la salida resultante suficiente para la descripción de la respuesta del sistema.

Otra base frecuentemente empleada corresponde a  $\phi(t, s) = e^{j2\pi st}$ ,  $T, S \subset (-\infty, \infty)$ . En este caso, se obtiene que  $v_x(s) = \mathcal{F}\{x(t)\} = X(s)$ , además,  $v_y(s) = \mathcal{F}\{y(t)\} = Y(s)$ . De acuerdo con (1.34), se tiene

$$\psi(t, s) = \int_T h(t, \tau) e^{j2\pi s\tau} d\tau \triangleq H(t, s) \quad (1.36)$$

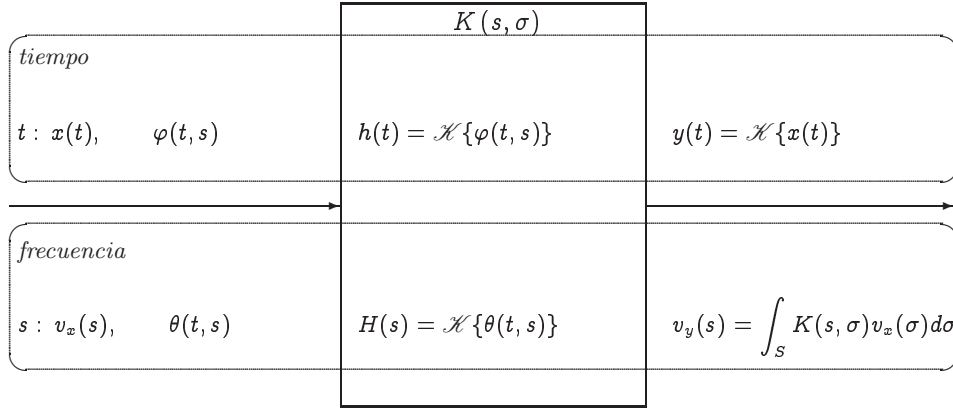
Por consiguiente, de (1.34) y (1.35) se obtiene

$$Y(f) = \int_{-\infty}^{\infty} H(f, \nu) X(\nu) d\nu \quad (1.37)$$

$$\text{donde } H(f, \nu) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h(t, \tau) e^{-j2\pi ft} e^{j2\pi \nu \tau} dt d\tau.$$

La respuesta de frecuencia  $H(f, \nu)$ , denominada *la función de transferencia*, corresponde a la segunda forma de análisis de sistemas conocida como el *método de análisis espectral*.

La relación entre la función de respuesta a impulso y la función de transferencia, así como con las respectivas bases de representación en la transformación lineal de señales se muestra en la Figura 1.3.



**Figura 1.3.** Transformación lineal e invariante de señales

Cabe anotar, que en algunos casos es importante representar las señales de entrada y salida con diferentes bases, así  $\varphi(t, s)$  puede corresponder a la entrada, mientras,  $\hat{\varphi}(t, s)$  a la salida, con lo cual la respectiva densidad de salida es:

$$v_y(s) = \int_s K(s, \sigma) v_x(\sigma) d\sigma$$

donde  $K(s, \sigma) = \int_T \hat{\theta}(s, t) \psi(t, \sigma) dt, \quad \psi(t, s) = \mathcal{K}\{\varphi(t, s)\}$

El tercer método de análisis de sistemas lineales se basa en su descripción generalizada mediante ecuaciones diferenciales, también lineales, pero de orden finito y conocido como el *método del operador diferencial de orden finito*,

$$\alpha_0(t) y(t) + \alpha_1(t) \frac{dy}{dt} + \cdots + \alpha_m(t) \frac{d^m y}{dt^m} = \beta_0(t) x(t) + \beta_1(t) \frac{dx}{dt} + \cdots + \beta_n(t) \frac{d^n x}{dt^n}, \quad (1.38)$$

siendo, respectivamente, las funciones continuas de entrada y salida  $\{\alpha_m(t)\}$  y  $\{\beta_n(t)\}$ .

**Sistemas lineales e invariantes en el tiempo.** La condición de invariabilidad, implica que el operador dependa sólo de la diferencia de los argumentos,  $y(t - t_0) = \mathcal{K}\{x(t - t_0)\}$ , que en los sistemas lineales, implica que la respuesta a impulso  $h(t, \tau)$  dependa solamente de la diferencia de los argumentos,  $h(t, \tau) = h(t - \tau)$ , con lo que (1.35) se convierte en

$$y(t) = \int_T h(t - \tau) x(\tau) d\tau \quad (1.39)$$

Así mismo, la descripción, dada en términos de la frecuencia lineal,  $f$ , para el operador

invariante en el tiempo, contiene el ajuste de los argumentos en (1.37):

$$\begin{aligned} H(f, \nu) &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h(t - \tau) e^{-j2\pi ft + j2\pi \nu \tau} dt d\tau = \int_{-\infty}^{\infty} h(\sigma) e^{-j2\pi f \sigma} \int_{-\infty}^{\infty} e^{j2\pi(f - \nu)\tau} d\tau d\sigma \\ &= H(f) \delta(f - \nu) \end{aligned}$$

por consiguiente

$$Y(f) = \int_{-\infty}^{\infty} H(f) \delta(f - \nu) X(\nu) d\nu = H(f) X(f) \quad (1.40)$$

Asumiendo  $x(t) = e^{j2\pi ft}$ , la convolución dada en (1.39) se define como:

$$\begin{aligned} y(t) &= \int_{-\infty}^{\infty} h(\tau) x(t - \tau) d\tau = \int_{-\infty}^{\infty} h(\tau) e^{j2\pi f(t - \tau)} d\tau = e^{j2\pi ft} \int_{-\infty}^{\infty} h(\tau) e^{-j2\pi f \tau} d\tau \\ &= e^{j2\pi ft} \mathcal{F}\{h(t)\} = e^{j2\pi ft} H(2\pi f) \end{aligned}$$

Luego, la función de transferencia  $H(2\pi f)$ , expresada en términos de frecuencia angular, corresponde a la TF de la función respuesta al impulso:

$$H(2\pi f) = \mathcal{F}\{h(t)\} = \int_{-\infty}^{\infty} h(t) e^{-j2\pi ft} dt$$

Por último, aplicando la TIF a (1.40), la salida se puede expresar en forma alterna:

$$y(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} H(2\pi) X(2\pi) e^{-j2\pi t} d\omega$$

El método del operador diferencial asume que en los sistemas invariantes en el tiempo los parámetros de entrada y salida en la ecuación diferencial (1.38) son constantes, con lo cual la correspondiente función respuesta a impulso de (1.39) se determina como [4]:

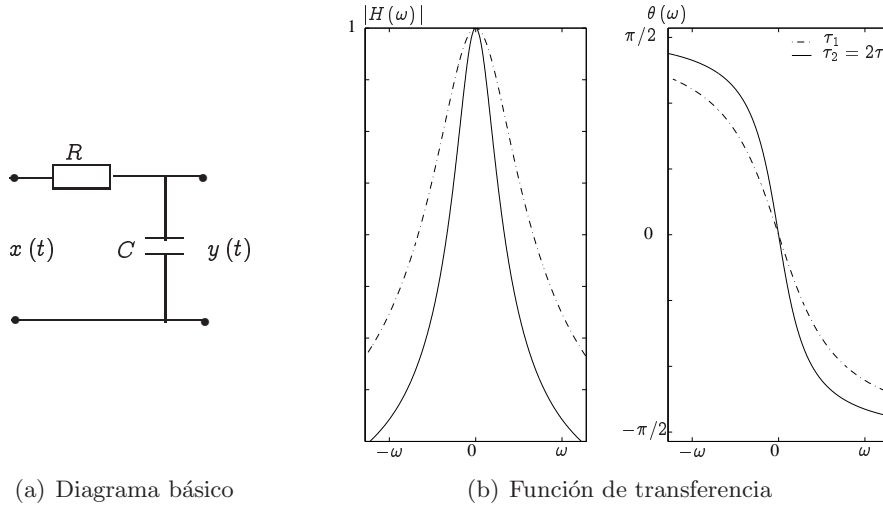
$$h(t - \tau) = \begin{cases} \sum_{k=1}^m \alpha_k e^{p_k(t - \tau)}, & t \geq \tau \\ 0, & t < \tau \end{cases} \quad (1.41)$$

donde  $p_k$  son las raíces del polinomio,  $Q(p) = \sum_{k=1}^m \alpha_k p^k = 0$ .

Al operador de (1.41) le corresponde la función de transferencia, que se expresa en la siguiente forma racional:

$$H(f) = \sum_{k=1}^m \frac{\alpha_k}{j\omega - p_k} = \frac{P(p)}{Q(p)}, \quad \alpha_k, \beta_k \in \mathbb{R}, \quad P(p) = \sum_{k=1}^n \beta_k p^k$$

**Ejemplo 1.3.** Determinar la función de transferencia del circuito  $RC$  en la Figura 1.4(a).



**Figura 1.4.** Análisis espectral del circuito  $RC$

Aplicando la relación del divisor de voltaje, se tiene:

$$y(t) = \frac{Z_c}{R + Z_c} e^{j\omega t}$$

donde la impedancia de la capacitancia es  $Z_c = 1/j\omega C$ . Así,

$$H(\omega) = \frac{(1/j\omega C)}{R + (1/j\omega C)} = \frac{1}{1 + j\tau\omega} \quad (1)$$

siendo  $\tau = RC$ . Al convertir en forma polar la expresión (1) se obtienen las respectivas *magnitud*,  $|H(\omega)| = \left(1 + (\omega\tau)^2\right)^{-1/2}$ , y *desfase*, que se muestran en la Figura 1.4(b).

#### 1.2.4. Representación de sistemas no lineales

El comportamiento de sistemas variantes en el tiempo, pero que trabajan en régimen no lineal, se describe por ecuaciones diferenciales también no lineales. La integración exacta de tales ecuaciones, en la mayoría de casos prácticos, es imposible. Actualmente, existe una serie de métodos de solución aproximada, los cuales básicamente, representan el dispositivo en análisis como una secuencia de conexiones lineales e invariantes sin retroalimentación. En este caso, se deben hacer los ajustes necesarios derivados de la aparición de nuevas componentes espectrales al paso de señales por sistemas no lineales. Aunque, cuando se tienen sistemas no lineales con lazos de retroalimentación (osciladores, sistemas de seguimiento, etc.), su análisis se hace mucho más complejo. Algunos de los métodos de representación corresponden a los descritos a continuación:



**Análisis del régimen de estado estable.** En este caso, se toma el modelo de estabilidad asintótica del sistema, aunque, de igual forma se puede asumir la variabilidad o invariabilidad de sus parámetros en el tiempo. El análisis puede realizarse tanto en el tiempo como en la frecuencia.

El método espectral se puede emplear cuando el sistema es abierto y consiste de una conexión en cascada de dispositivos lineales invariantes o no lineales variantes en el tiempo. En primera instancia, puede considerarse el modelo de un dispositivo no lineal con una función de transformación entrada-salida, descrita por  $y(t) = \mathcal{K}\{x(t)\}$ , a cuya salida le corresponde la TF:

$$Y(\omega) = \int_{-\infty}^{\infty} \mathcal{K}\{x(t)\} e^{-j\omega t} dt \quad (1.42)$$

que determina de forma analítica la función que aproxima la característica del dispositivo no lineal.

En general, la señal de entrada se puede representar mediante su descomposición espectral, por ejemplo, mediante la serie de Fourier, luego se obtiene para cada una de las componentes, su respectiva salida después de cada dispositivo, teniendo en cuenta que la señal total de salida superpone todas las respuestas elementales del sistema:

$$y = \mathcal{K}\left\{\sum_{k=1}^n x_n e^{jk\omega_0 t}\right\}$$

La segunda forma de representación de un sistema, en régimen de estado estable, corresponde a su desarrollo en el tiempo. Si se asume la invariabilidad del sistema, todas las derivadas en el tiempo de los respectivos parámetros deben ser cero, con lo cual se puede analizar el régimen de estado estable con base en la respectiva ecuación diferencial, que describe el sistema, la cual después de hacerse cero se convierte en una ecuación algebraica simple. La solución de esta clase de ecuaciones se puede realizar por métodos gráficos o de aproximaciones sucesivas numéricas [6].

**Ejemplo 1.4.** Hallar la DEE de la corriente de salida relacionada con el voltaje de entrada mediante la expresión,

$$i(t) = \begin{cases} k(u_1 e^{-\alpha t} + u_0) > 0, & t_1 \leq t \leq t_2 \\ 0, & \text{otros valores de } t \end{cases}$$

donde  $k$  es una constante.

La densidad espectral de energía del voltaje de entrada, teniendo en cuenta el valor del retardo  $t_1$  tiene la forma

$$U(\omega) = \int_{t_1}^{\infty} u_1 e^{-\alpha(t-t_1)} e^{-j\omega t} dt = u_1 \frac{e^{j\omega t_1}}{\alpha + j\omega}$$

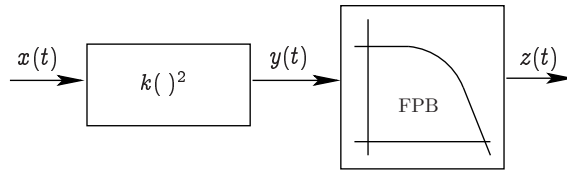
mientras a la salida, a partir de (1.42), le corresponde la densidad espectral,

$$I(\omega) = \int_{t_1}^{t_2} k \left( u_1 e^{-\alpha(t-t_1)} + u_0 \right) e^{-j\omega t} dt = k e^{-j\omega t_1} \int_0^{t_2-t_1} (u_1 e^{-\alpha t} + u_0) e^{-j\omega t} dt$$

$$= k u_1 \frac{e^{-j\omega t_1}}{\alpha + j\omega} \frac{\alpha e^{-(\alpha+j\omega)(t_2-t_1)} - (\alpha + j\omega) e^{-\alpha(t_2-t_1)} + j\omega}{j\omega}$$

cuya forma se hace compleja, como resultado de la no linealidad del proceso.

**Ejemplo 1.5.** Analizar el régimen estable de una señal compuesta de dos componentes armónicos:  $x(t) = x_1 \cos(\omega_1 t) + x_2 \cos(\omega_2 t)$ , a su paso el sistema mostrado en la Figura 1.5, compuesto por un elevador al cuadrado (con función entrada-salida  $y(t) = kx^2(t)$ ) y un filtro  $RC$  pasabajos, como el descrito por la función de transferencia (1) del ejemplo 1.3.



**Figura 1.5.** Detector cuadrático

A la salida del elevador al cuadrado se tiene

$$\begin{aligned} y(t) &= kx^2(t) = k(x_1 \cos(\omega_1 t) + x_2 \cos(\omega_2 t))^2 \\ &= kx_1^2 \cos^2(\omega_1 t) + 2kx_1 x_2 \cos(\omega_1 t) \cos(\omega_2 t) + kx_2^2 \cos^2(\omega_2 t) \\ &= \frac{k}{2} (x_1^2 + x_2^2) + \frac{k}{2} x_1^2 \cos(2\omega_1 t) + \frac{k}{2} x_2^2 \cos(2\omega_2 t) + kx_1 x_2 (\cos(\omega_3 t) + \cos(\omega_4 t)) \end{aligned}$$

donde  $\omega_3 = \omega_1 - \omega_2$  y  $\omega_4 = \omega_1 + \omega_2$ .

Teniendo en cuenta las expresiones obtenidas para la magnitud y desfase del filtro  $RC$  en el ejemplo 1.3, la salida del detector cuadrático tiene la forma,

$$\begin{aligned} z(t) &= \frac{k}{2} (x_1^2 + x_2^2) + \frac{k}{2} \frac{x_1^2}{\sqrt{1 + 4\omega_1^2 \tau^2}} \cos(2\omega_1 t + \theta_1) + \frac{k}{2} \frac{x_2^2}{\sqrt{1 + 4\omega_2^2 \tau^2}} \cos(2\omega_2 t + \theta_2) + \\ &+ k \frac{x_1 x_2}{\sqrt{1 + \omega_3^2 \tau^2}} \cos(\omega_3 t + \theta_3) + k \frac{x_1 x_2}{\sqrt{1 + \omega_4^2 \tau^2}} \cos(\omega_4 t + \theta_4) \end{aligned}$$

donde  $\theta_{1,2} = -\arctan(2\omega_{1,2}\tau)$ ,  $\theta_{3,4} = -\arctan(\omega_1 \pm \omega_2\tau)$ . Si se asume un valor suficientemente grande la constante  $\tau$ , la salida del filtro pasabajos (FPB) se puede aproximar al siguiente valor:

$$z(t) \approx \frac{k}{2} (x_1^2 + x_2^2)$$

**Método de series funcionales.** También conocido como el *método de Wiener* [7], puede ser catalogado como heurístico y se fundamenta en la posibilidad de representar sistemas no lineales e invariantes en el tiempo en forma de series.

Sea un sistema con entrada  $x(t)$  y salida  $y(t)$ ; ambas relacionadas por el operador  $\mathcal{K}$ :

$$y(t) = \mathcal{K}\{x(t)\}$$

En general, el operador  $\mathcal{K}$  depende del carácter del sistema, su objeto de análisis y tipo de entradas. En el caso concreto, se escoge una clase de sistemas con memoria finita, asumiendo que el operador es biyectivo. Esta restricción no aplica en sistemas en los cuales existan puntos de equilibrio dependientes de las condiciones iniciales de entrada (sistemas con memoria infinita).

Una forma de representación de la no linealidad en un sistema, más angosta que las correspondientes ecuaciones diferenciales, corresponde a las *series de Volterra* [8]:

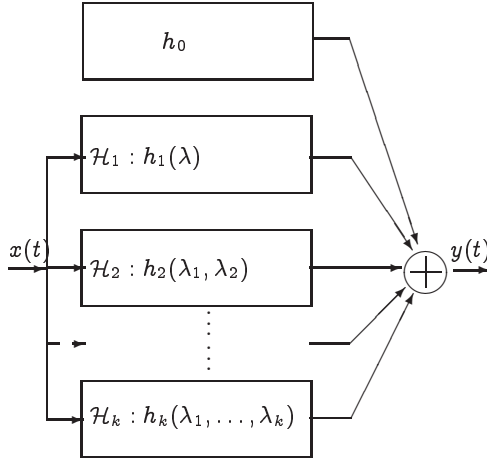
$$y(t) = h_0(t) + \sum_k \int \cdots \int h_k(t, \tau_1, \dots, \tau_k) x(\lambda_1) \cdots x(\lambda_k) d\lambda_1 \cdots d\lambda_k \quad (1.43)$$

En particular, cuando se tienen ecuaciones diferenciales de primer orden [9],

$$\frac{dy}{dt} = f(t, y(t)) + g(t, y(t), x(t))$$

la representación (1.43) es viable, por ejemplo, cuando la función  $f(t, y(t))$  con respecto a  $y$  se puede aproximar por la serie de Taylor, mientras la función  $g$  no dependa de  $y(t)$ .

Con el objeto de simplificar el análisis, el modelo de representación se puede limitar hasta los sistemas invariantes en el tiempo, cuando se asume que es cierta la propiedad  $\mathcal{K}\{x(t+\tau)\} = y(t+\tau)$ ,  $\forall \tau$ , por esta razón, la serie funcional de Volterra (1.43) toma una forma más simple,



**Figura 1.6.** Representación por series de Volterra

$$y(t) = h_0 + \int_{-\infty}^{\infty} h_1(\lambda_1) x(t - \lambda_1) d\lambda_1 + \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h_2(\lambda_1, \lambda_2) x(t - \lambda_1) x(t - \lambda_2) d\lambda_1 d\lambda_2 + \cdots \quad (1.44)$$

Las funciones  $h_k(\lambda_1, \dots, \lambda_k)$  son denominadas los *núcleos de Volterra*, que al igual que en los sistemas lineales, representan la respuesta a la función  $\delta(t)$ . El primer término  $h_0$  se puede tomar como componente constante de la representación, o bien, como las condiciones iniciales, por cuanto  $y(t) = h_0$ , cuando  $x(t) = 0$ . En cuanto al segundo término,

$\int h_1(\lambda_1) x(t - \lambda_1) d\lambda_1$ , si cumple la condición de realización física,  $h_1(t) = 0, \forall t < 0$ , entonces éste corresponde a la parte lineal del sistema.

El tercer término,  $\int \int h_2(\lambda_1, \lambda_2) x(t - \lambda_1) x(t - \lambda_2) d\lambda_1 d\lambda_2$ , corresponde a la convolución doble de la señal de entrada  $x(t)$  con la respuesta a impulso  $h_2(\lambda_1, \lambda_2)$ . Si se asume la independencia de los términos  $h_2(\lambda_1, \lambda_2) = h(\lambda_1)h(\lambda_2)$ , entonces el tercer término es:

$$\begin{aligned} \int \int h_2(\lambda_1, \lambda_2) x(t - \lambda_1) x(t - \lambda_2) d\lambda_1 d\lambda_2 \\ = \int h(\lambda_1) x(t - \lambda_1) d\lambda_1 \int h(\lambda_2) x(t - \lambda_2) d\lambda_2 = y(t) y(t) = y^2(t) \end{aligned}$$

esto es, el tercer término puede ser entendido como la descripción de un sistema cuadrático.

En general, asumiendo que todos los núcleos de Volterra cumplen las condiciones:

- *Realización física.*  $h_k(\lambda_1, \dots, \lambda_n) = 0, \forall \tau_k < 0, k = 1, \dots, n$
- *Simetría.*  $h_k(\lambda_1, \dots, \lambda_n) = h_k(\lambda_n, \dots, \lambda_1) = \dots$ , esto es, el núcleo es el mismo cualquiera que sea el orden de sus argumentos.

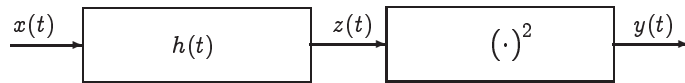
entonces, un sistema no lineal dado con memoria finita puede ser representado por la serie de Volterra (1.44), en la cual cada término puede ser asociado con su respectivo operador elemental  $\mathcal{H}_k$ , como se muestra en la Figura 1.6:

$$y(t) = h_0 + \sum_{k \geq 0} \mathcal{H}_k \{x(t)\} \quad (1.45)$$

donde  $\mathcal{H}_k \{x(t)\} = \int \dots \int h_k(\lambda_1, \dots, \lambda_k) x(\lambda_1) \dots x(\lambda_k) d\lambda_1 \dots d\lambda_k$ .

---

**Ejemplo 1.6.** Analizar la convergencia del sistema invariable en el tiempo, descrito por la función linealidad  $y(t) = z(t) (1 + z^2(t))^{-1}$ , representado en la Figura 1.7.



**Figura 1.7.** Sistema cuadrático simple

La descomposición de la función, mediante análisis en la serie de Taylor, tiene la forma:

$$y(t) = \sum_{k=0}^{\infty} (-1)^k (z(t))^{2k+1} \quad (1)$$

Reemplazando, por el núcleo básico de primer orden  $z(t) = h(t) * x(t)$ , se obtiene la respectiva serie de Volterra

$$y(t) = \sum_{k=0}^{\infty} (-1)^k \left( \int_{-\infty}^{\infty} h(\lambda) x(t - \lambda) d\lambda \right)^{2k+1} = \sum_{k=0}^{\infty} \mathcal{H}_{2k+1} \{x(t)\} \quad (2)$$

donde los núcleos son de la forma  $h_{2k+1}(\lambda_1, \dots, \lambda_{2k+1}) = (-1)^k h(\lambda_1) \cdots h(\lambda_{2k+1})$ .

Por cuanto la serie de Taylor (1), obtenida para el sistema cuadrático simple, converge solo para valores de  $z(t) < 1$ , entonces la respectiva serie (2) converge en aquellos momentos del tiempo cuando  $|z(t)| \geq 1$ . Lo anterior implica, que en el ejemplo del sistema cuadrático simple, la serie de Volterra se puede emplear para señales de entrada, cuya salida  $z(t)$  no exceda el valor de 1.

---

Cabe anotar que la convergencia de la serie de Taylor se exige para todo el dominio de representación  $t$ . Así por ejemplo, si en la Figura 1.7, en vez del elevador a cuadrado, se tiene un limitador bilateral del tipo  $y(t) = a \operatorname{sgn}(z(t))$ , entonces para esta función con rompimiento en el punto  $t = 0$ , la serie de Taylor no existe en ese mismo punto, y por lo tanto, tampoco existe la representación en series de Volterra.

## Problemas

**Problema 1.7.** Demostrar que el siguiente conjunto de señales es ortogonal y calcular su TF directa:

$$\Phi_n(t) = \text{sinc } \Omega(t - nT), \quad n = 0, \pm 1, \pm 2, \dots, -\infty < t < \infty.$$

**Problema 1.8.** Calcular los coeficientes de la serie de Fourier y la TF de la función dada en la forma,  $x(t) = \sum_{n=-\infty}^{\infty} \delta(t - nT)$ .

**Problema 1.9.** Hallar la TF de un pulso Gaussiano  $x(t) = a \exp -\pi(t/\tau)^2$  y verificar el efecto de extensión recíproca de escala para  $\tau = 1$  y  $\tau = 2$ . Calcular la diferencia de las DEE en los puntos de intersección.

**Problema 1.10.** Las señales  $x_p(t) = x_p(-t)$  y  $x_i(t) = -x_i(-t)$  se relacionan con la función  $y(t)$  como:

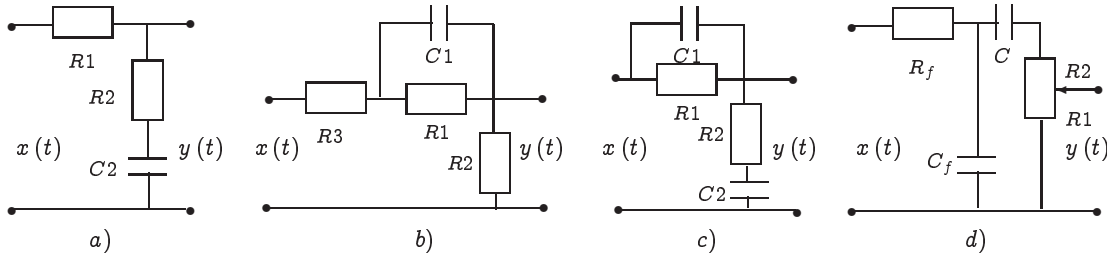
$$x_p(t) = y(t) + y(-t)$$

$$x_i(t) = y(t) - y(-t)$$

Hallar la relación entre las respectivas DEE  $X_k(\omega) = \mathcal{F}\{x_k(t)\}$ ,  $k \in \{p, i\}$  con la densidad  $Y(\omega)$ .

**Problema 1.11.** Sea  $X(\omega) = \mathcal{F}\{x(t)\}$ . Hallar la señal  $y(t) = \mathcal{F}^{-1}\{Y(\omega)\}$ , para cada uno de los siguientes casos: a)  $Y(\omega) = X^2(\omega)$ , b)  $Y(\omega) = X(\omega)X^*(\omega)$  y c)  $Y(\omega) = X^*(\omega)$ .

**Problema 1.12.** Conformar las ecuaciones diferenciales de los circuitos dados en la Figura 1.8.



**Figura 1.8.** Circuitos de filtración

**Problema 1.13.** Determinar la respuesta a impulso para cada uno de los circuitos descritos por las siguientes ecuaciones diferenciales:

$$\begin{aligned} T \frac{dy}{dt} - y &= kx & y &= k \left( x + T \frac{dx}{dt} \right) \\ y &= k \left( x + 2k_1 T \frac{dx}{dt} + T^2 \frac{d^2x}{dt^2} \right) & T^2 \frac{d^2y}{dt^2} + 2k_1 T \frac{dy}{dt} + y &= kx \\ T^2 \frac{d^2y}{dt^2} - 2k_1 T \frac{dy}{dt} + y &= kx & \frac{1}{\omega^2} \frac{d^2y}{dt^2} + y &= kx \end{aligned}$$

**Problema 1.14.** Comparar los espectros de entrada y salida del dispositivo no lineal con tensión de salida  $i = a_0 + a_1(u - u_0) + a_2(u - u_0)^2 + a_3(u - u_0)^3$ , cuando el voltaje de entrada tiene la forma  $u = u_1(t) + u_0$ , asumiendo que la componente variable corresponde a un pulso triangular,

$$u_1(t) = \begin{cases} bt, & 0 \leq t \leq T \\ 0 & \text{otros valores de } t \end{cases}$$

## 1.3. Discretización de señales y sistemas

### 1.3.1. Discretización uniforme

El empleo de sistemas discretos en campos donde las señales son continuas supone su previa acomodación, esto es, la *discretización* de las señales. El modelo conveniente para la discretización ideal, o muestreo de una señal continua  $x_c(t)$ , está dado por su multiplicación con alguna señal periódica de discretización  $x_d(t)$ , con el fin de obtener la señal discretizada  $x(k\Delta t)$ :

$$x(k\Delta t) = x_c(t)x_d(t) = \sum_k x_c(k\Delta t)\delta(t - k\Delta t) \quad (1.46)$$

donde  $x_d(t) = \sum_{k=-\infty}^{\infty} \delta(t - k\Delta t)$ , siendo  $\Delta t$  el período de discretización.

A partir de la expresión (1.46) se observa que la señal discretizada  $x(t)$  se define solamente en los momentos equidistantes de tiempo  $t = k\Delta t$ , luego entonces, parte de la información de la señal original  $x_c(t)$  se pierde durante su discretización.

**Teorema 1.3.** (*Discretización de Kotelnikov.*) Sea una señal continua  $x_c(t)$ , para la cual no se consideran componentes espectrales mayores a  $\omega_{\max} = 2\pi f_{\max}[\text{rad/s}]$ , donde  $f_{\max}$  es el máximo valor de frecuencia para la cual  $X(\omega) \neq 0$ , entonces, toda la información de la señal continua estará enteramente contenida en los valores  $x_c(n\Delta t)$ , asegurando que se cumpla la desigualdad:

$$\Delta t \leq 1/2f_{\max} \quad (1.47)$$

El teorema 1.3 permite representar la señal continua  $x_c(t)$  en forma de la serie:

$$x_c(t) = \sum_{k=-\infty}^{\infty} x_c(k\Delta t) \frac{\sin(\omega_{\max}(t - k\Delta t))}{\omega_{\max}(t - k\Delta t)} \quad (1.48)$$

Al comparar la serie (1.48) con la representación (1.4), se observa que las funciones base de descomposición son las siguientes:

$$\phi_k(t) = \frac{\sin(\omega_{\max}(t - k\Delta t))}{\omega_{\max}(t - k\Delta t)}$$

La serie de Kotelnikov (1.48) permite el restablecimiento de la función inicial  $x_c(t)$ , en cualquier momento del tiempo  $t$  (dentro del intervalo de análisis). Sin embargo, el restablecimiento de  $x_c(t)$  exige realizar la sumatoria sobre una cantidad infinita de términos, lo cual prácticamente es imposible. De otra manera, la implementación del teorema de discretización conlleva inevitablemente a errores de representación.

Cabe anotar, que los valores sucesivos de los tiempos discretización uniforme, descritos por la malla  $\{n\Delta t : n \in \mathbb{Z}\}$ , están separados a un intervalo constante e igual a  $\Delta t$ , cumpliendo (1.47), por lo que en la práctica es común normalizar su valor, esto es,

$$x_c(n\Delta t) = x_d[n]$$

### 1.3.2. Transformadas ortogonales discretas

La realización de cualquier forma de representación integral de señales mediante sistemas reales de proceso digital, dada una señal discreta  $x[n]$ , se encuentra con algunas dificultades insuperables. En particular, se necesitan procesadores con memoria infinita para guardar, tanto la infinita cantidad de valores de la serie de entrada  $x[n]$ , como el conjunto de valores del continuo de la respectiva representación espectral  $H(e^{j\Omega})$ . En consecuencia, empleando el modelo de las respectivas bases continuas, en la representación generalizada de Fourier en (1.12), y para efectos de realización práctica, el conjunto de transformadas ortogonales discretas se define estrictamente sobre sucesiones discretas,  $x[n]$ , con longitud finita  $N$ , o bien sobre sucesiones periódicas con periodo  $N$ , de la siguiente forma:

$$x[k] = \sum_{n=0}^{N-1} v[n] \phi_k[n], \quad k = 0, 1, \dots, N-1 \quad (1.49a)$$

$$v[n] = \frac{1}{N} \sum_{k=0}^{N-1} x[k] \phi_n^*[k], \quad n = 0, 1, \dots, N-1 \quad (1.49b)$$

Entre los diferentes conjuntos ortogonales analizados en la sección §1.1.3, en el proceso digital de señales moderno, encuentran aplicación las transformadas discretas de Fourier (con sus respectivos casos particulares de la T. *Hartley* y la T. *discreta de cosenos*), de *Walsh* y *Haar*, entre otras. En el caso de la Transformada discreta de Fourier (TDF), en calidad de sistema base se emplea la base ortogonal exponencial (1.8), pero en su forma discretizada:

$$\phi_k^*(n) = e^{jkn2\pi/N} \triangleq W_N^{kn}, \quad n = 0, 1, \dots, N-1 \quad (1.50)$$

El *factor de pivote*  $W_N = e^{j2\pi/N}$  cumple la condición de ortogonalidad:

$$\sum_{n=0}^{N-1} W_N^{kn} W_N^{-mn} = \begin{cases} N, & (k-m) = lN, l \in \mathbb{N} \\ 0, & (k-m) \neq lN \end{cases} \quad (1.51)$$

Así mismo, este factor es periódico con período  $N$ , tanto en el sentido de la variable de frecuencia  $n$ , como el de la variable de tiempo  $k$ :

$$W_N^{kn} = W_N^{(k+N)n} = W_N^{k(n+N)} \quad (1.52)$$

Teniendo en cuenta las ecuaciones (1.50), (1.51) y (1.52), se determina el par de transformadas discretas de Fourier, en función del factor de pivote, de la siguiente forma:

$$x[k] = \sum_{n=0}^{N-1} X[n] W_N^{-kn}, \quad k = 0, 1, \dots, N-1 \quad (1.53a)$$

$$X[n] = \frac{1}{N} \sum_{k=0}^{N-1} x[k] W_N^{kn}, \quad n = 0, 1, \dots, N-1 \quad (1.53b)$$



El sistema de transformadas (1.53a) y (1.53b) establecen la relación mutua e unívoca entre los valores de la serie de entrada  $x[k]$  y su representación discreta espectral  $X[n]$ . Por cuanto el factor de pivote es periódico, entonces el espectro discreto de Fourier de la señal también es periódico, esto es,  $X(n) = X(n + mN)$ ,  $m = \pm 1, \pm 2, \dots$

### 1.3.3. Representación ortogonal de sistemas en tiempo discreto

**Respuesta discreta a impulso.** Los sistemas lineales en tiempo discreto se describen por medio de ecuaciones de diferencias con parámetros constantes:

$$\sum_{n=0}^N a_n y[k - n] = \sum_{n=0}^M b_n x[k - n] \quad (1.54)$$

donde  $a_0 = 1$ , los términos  $y[k - n]$  y  $x[k - n]$  corresponden a los valores de las señales discretizadas de salida y entrada, respectivamente. Si se cumple que  $a_n \neq 0, \forall n \neq 0$ , entonces el sistema se denomina *recursivo*. La señal a la salida de tal sistema depende no solamente de las señales de entrada, sino de valores de la señal de salida en momentos anteriores. En cambio, cuando se tiene que los coeficientes  $a_n = 0, n = 1, 2, \dots, N$ , entonces, se habla de un sistema *no recursivo*, cuya señal de salida se determina solamente por la acción de entrada hasta el momento actual de tiempo.

La respuesta a impulso de un sistema discreto  $h[n]$  se define como la salida cuando la entrada corresponde a la función  $x[n] = \delta[n]$ . La señal a la salida de un sistema discreto lineal e invariante en el tiempo se define como:

$$y[n] = \sum_{k=-\infty}^{\infty} h[n, k] x[k] = \sum_{k=-\infty}^{\infty} h[n - k] x[k], \quad (1.55)$$

Asumiendo la condición de realización física del sistema, que implica  $h[n, k] = 0, \forall k > n$ , entonces, se obtiene que,  $y[n] = \sum_{k=0}^{\infty} h[k] x[n - k], n = 0, 1, \dots$

En los filtros no recursivos, la respuesta a impulso contiene una cantidad finita de términos, que no excede el valor de  $M + 1$ , donde  $M$  es el límite superior de la suma en (1.55), correspondiente a los elementos de la entrada; estos sistemas se denominan de *respuesta a impulso finita*:

$$h_{RIF}[n] = \sum_{i=n_0}^N \alpha_i \delta[n - i], \quad \forall n_0, N < \infty \quad (1.56)$$

Por el contrario, los sistemas discretos recursivos que poseen una respuesta a impulso infinita se denominan de *respuesta a impulso infinita*:

$$h_{RII} = \sum_{i=n_0}^N \alpha_i \delta[n - i], \quad |N - n_0| \rightarrow \infty \quad (1.57)$$

**Función de transferencia discreta.** Como en caso continuo, en los sistemas discretos lineales e invariantes en el tiempo con respuesta a impulso dada  $h[n]$ , se determina la señal de salida  $y[n]$ , asumiendo que se tiene como señal de entrada  $x[n] = \exp[jn\Omega]$ , esto es,

$$y[n] = \sum_{i=-\infty}^{\infty} h[i]x[n-i] = \sum_{i=-\infty}^{\infty} h[i]e^{jn\Omega} = e^{jn\Omega} \sum_{i=-\infty}^{\infty} h[i]e^{-ji\Omega}$$

teniendo en cuenta (1.36), entonces  $y[n] = e^{jn\Omega} H(e^{j\Omega})$ , donde  $H(e^{j\Omega})$  es la respuesta de frecuencia o *función de transferencia discreta*, que en forma general, puede ser compleja como se demuestra en el caso de los sistemas continuos, así,

$$Y(e^{j\Omega}) = X(e^{j\Omega}) H(e^{j\Omega}),$$

entonces,

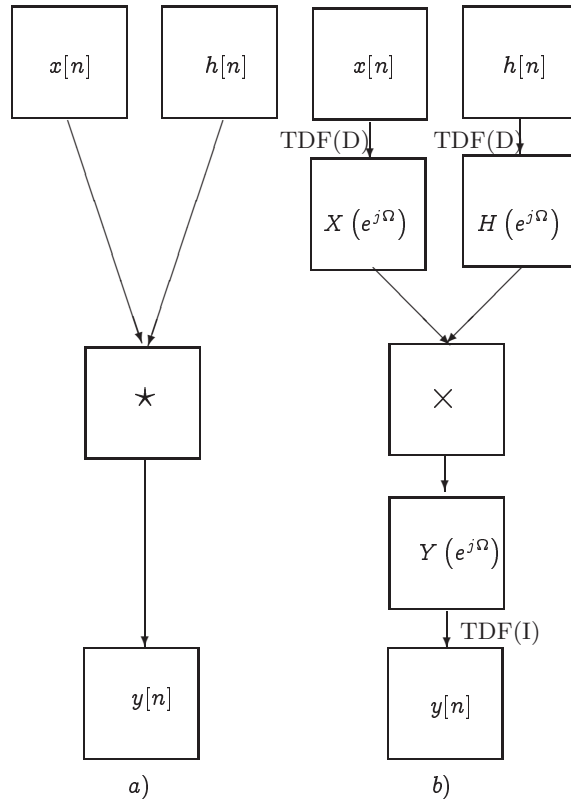
$$y[n] = x[n] * h[n]$$

Los métodos de análisis de sistemas discretos, básicamente, corresponden a los mismos métodos del caso continuo, los cuales se representan en la Figura 1.9.

De acuerdo con el desarrollo matemático anteriormente descrito de la función de transferencia, para su determinación existen varios métodos. El primer método emplea la función de impulso unitario  $\delta[n]$  como señal de entrada, un segundo método emplea la función exponencial,  $e^{jn\Omega}$ , en correspondencia con la misma definición y, por último, sin especificar la señal de entrada se puede analizar la ecuación de diferencias que concretamente describe el sistema [10].

De otra parte, el método matricial es común en la descripción de sistemas discretos, cuando el operador  $\mathcal{H}$  se representa en forma de matriz. Sean, en calidad de funciones propias de un operador, las siguientes funciones discretas  $\phi_m[k]$ ,  $m, k = 0, 1, \dots, N-1$ , esto es,

$$\mathcal{H}\{\phi_m[k]\} = H[m]\phi_m[k],$$



**Figura 1.9.** Análisis de sistemas discretos

La filtración discreta se puede entender como la multiplicación de los coeficientes  $v[n]$ ,

obtenidos de la representación de la señal de entrada con la base  $\phi_m[k]$ , por los respectivos coeficientes de la función de transferencia  $H[m]$ . La multiplicación corresponde a los coeficientes espectrales de la señal de entrada.

Las restricciones de implementación práctica hacen que en el proceso de señales se utilicen solamente bases discretas del tipo ortogonal, para los cuales la señal de salida puede ser representada por la siguiente expresión matricial:

$$\mathbf{Y}_{N \times 1} = \frac{1}{N} \Phi_N^H \mathbf{H}_N \Phi_N \mathbf{X}_N$$

donde  $\mathbf{Y}$  y  $\mathbf{X}$  se forman de las respectivas señales discretizadas de salida y entrada,  $\Phi$  es una matriz ortogonal, con filas determinadas por las funciones discretas ortogonales  $\phi_m[k]$ ,  $m, k = 0, \dots, N-1$ , donde  $N$  es el conjunto de la base y  $\mathbf{H}$  es la matriz ortogonal compuesta por los coeficientes  $H[m]$ . De lo anterior, se deduce que la transformación lineal discreta se representa por la matriz

$$\mathcal{H} = \frac{1}{N} \Phi^H \mathbf{H} \Phi \tag{1.58}$$

## Problemas

**Problema 1.15.** Sea la señal de video  $x(t) = x_0 \exp(-\alpha t) u(t)$ . Seleccionar el valor del intervalo de muestreo  $\Delta$ , de tal manera que la magnitud de la DEE en la frecuencia límite  $\omega_{\max}$  se disminuye hasta el valor de  $0.01X(0)$ .

**Ejercicio en el CP 1.2.** Sea una señal seno  $x(t)$  definida en el intervalo de tiempo  $[0, 1]$ . Hallar su representación discretizada a una velocidad de muestreo 0.01.

**Ejercicio en el CP 1.3.** Dado el siguiente código, analizar las siguientes propiedades del factor de pivote:

- periodicidad, (1.52),
- antisimetría

```
N=2,3,4,5,6,7,8; % Escala de tiempo
for n = 1:5*N
    m(n)=exp(-j*2*pi*n/N);
end
m % valores del factor de pivote
```

**Problema 1.16.** Calcular la convolución lineal de las siguientes sucesiones finitas:

$$\mathbf{x} = [2 \ -2 \ 1]^T, \mathbf{h} = [1 \ 2]^T. \quad \mathbf{x} = [2 \ -2 \ 1]^T, \mathbf{h} = [1 \ 2 \ 0 \ 0]^T.$$

**Problema 1.17.** Dado un filtro de respuesta a impulso finita, para un valor de  $N = 9$ , que se aproxima a un FPB ideal con frecuencia de corte  $\Omega_c = 0.2\pi$  y respuesta a impulso deseada  $h[n] = \sin(0.2\pi n)/(\pi n)$ . Hallar la respectiva función de transferencia para los casos de ventanas: a) rectangular, b) Hamming.

**Problema 1.18.** Sea un circuito diferenciador ideal (transformador de Hilbert), para un valor de  $N = 15$ , con respuesta a impulso deseada

$$h[n] = \begin{cases} 0, & n \text{ par} \\ \frac{2}{n\pi}, & n \text{ impar} \end{cases}$$

Hallar la respectiva función de transferencia para los siguientes casos de ventana:

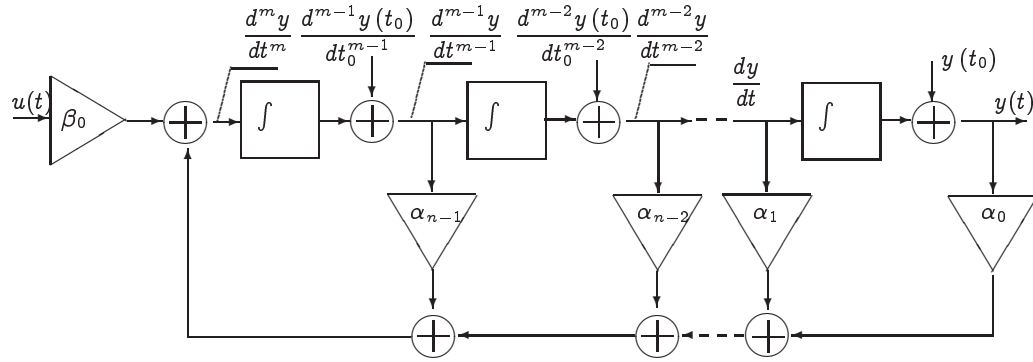
a) rectangular, b) Hamming.

**Problema 1.19.** Hallar las funciones de transferencia de los circuitos, descritos en los problemas 1.12 y 1.13.

## 1.4. Representación dinámica de sistemas

### 1.4.1. Sistemas lineales en variables de estado

A partir del método de ecuaciones diferenciales (1.38), que describe un sistema lineal con una entrada y una salida, en el numeral §1.2.3 se analiza la función de respuesta a impulso  $h(t)$  como forma de descripción del sistema. Sin embargo, el mismo método de ecuaciones diferenciales puede emplearse directamente en la representación de los sistemas dinámicos lineales. En este segundo caso, se deben tener en cuenta las siguientes condiciones:



**Figura 1.10.** Modelo de un sistema lineal de orden  $m$ .

- Las condiciones iniciales del sistema deben ser conocidas para determinar su respuesta en cualquier intervalo de tiempo  $t_0 \leq t \leq t_1$ . Además, deben estar en un número tal que describan completamente la influencia de la señal de entrada, hasta el instante  $t_0$ , sobre la señal de salida después del momento  $t_0$ .

El *estado del sistema* se define como la cantidad mínima de elementos de referencia del sistema necesaria para la descripción completa de su señal de salida, mientras las mismas variables de referencia se denominan *variables de estado*. En la práctica, el estado del sistema se describe por un vector de orden finito y, por lo tanto, se habla de un *sistema dinámico con dimensión finita*.

- La solución de la ecuación diferencial se puede implementar mediante los siguientes dispositivos de operación básica: integradores, sumadores, dispositivos no lineales, amplificadores de ganancia variable, etc. La conexión de estos dispositivos de operaciones básicas se considera el *modelo del sistema dinámico*.

La forma generalizada de una ecuación diferencial de orden  $m$ ,

$$\beta_0 u(t) = \alpha_0 y(t) + \alpha_1 \frac{dy}{dt} + \dots + \alpha_m \frac{d^m y}{dt^m}, \quad (1.59)$$

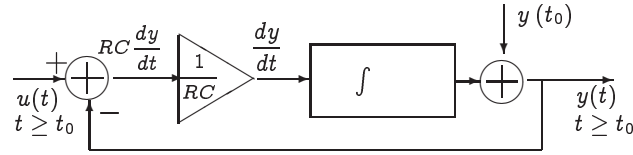
implica la selección de  $y(t), \dots, d^{m-1}y/dt^{m-1}$  en calidad de variables de estado, cuya

su solución se puede modelar empleando un procesador analógico equivalente como el mostrado en la Figura 1.10.

**Ejemplo 1.7.** Sea el circuito  $RC$  representado en la Figura 1.3, en el cual la señales de salida  $y(t)$  y entrada  $u(t)$  están relacionadas por la ecuación diferencial

$$RC \frac{dy}{dt} + y(t) = u(t) \quad (1.60)$$

El cálculo de la señal de salida (voltaje) en el intervalo  $t \geq t_0$ , implica el conocimiento de la señal de entrada para  $t \geq t_0$ , además del voltaje inicial  $y(t_0)$ , presente en el condensador en el momento  $t_0$ . De esta manera, es suficiente introducir una sola variable de estado, que es la que corresponde al voltaje  $y(t)$ . Así mismo, la expresión (1.60) es la ecuación de estado del circuito  $RC$  en análisis, en concordancia con la cual, se representa el respectivo modelo equivalente en la Figura 1.11.



**Figura 1.11.** Modelo equivalente del circuito  $RC$

Un sistema, dado por la ecuación (1.59) de orden  $m$ , puede ser descrito por ecuaciones diferenciales de primer orden. Frecuentemente, tal sistema de ecuaciones diferenciales, que es denominado *ecuación vectorial diferencial de primer orden*, es más fácil de analizar que la ecuación inicial de orden  $m$ . Así por ejemplo, en vez de la ecuación (1.59) se puede analizar el siguiente sistema:

$$\begin{aligned} x_1(t) &= y(t) \\ x_2(t) &= \frac{dy}{dt} = \frac{dx_1}{dt} \\ x_3(t) &= \frac{d^2y}{dt^2} = \frac{dx_2}{dt} \\ &\dots = \dots \\ x_n(t) &= \frac{d^{m-1}y}{dt^{m-1}} = \frac{dx_{m-1}}{dt} \end{aligned}$$

por lo tanto, se tiene que

$$\frac{du^m}{dt} = \frac{d^m y}{dt^m} = \beta_0 u(t) - \sum_{k=1}^m \alpha_{k-1} \frac{d^{k-1} y}{dt^{k-1}} = \beta_0 u(t) - \sum_{k=1}^m \alpha_{k-1} x_k(t), \quad t \geq t_0 \quad (1.61)$$

Si se introduce el vector columna de funciones  $\mathbf{x}(t) \in \{x_k : k = 1, \dots, m\}$ , para  $t \geq t_0$ , entonces en vez de la ecuación (1.59) de orden  $m$ , se analiza la siguiente ecuación equivalente

de primer orden y con dimensión  $m$ ,

$$\frac{d\mathbf{x}}{dt} = \mathbf{A}\mathbf{x}(t) + \mathbf{B}u(t) \quad (1.62)$$

donde las matrices  $\mathbf{A}_{m \times m}$  y  $\mathbf{B}_{n \times 1}$  tienen la forma

$$\mathbf{A} = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & \cdots & 1 \\ -\alpha_m & -\alpha_{m-1} & -\alpha_{m-2} & \cdots & -\alpha_1 \end{bmatrix}, \quad \mathbf{B} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ \beta_0 \end{bmatrix} \quad (1.63)$$

cuya solución requiere el vector de condiciones iniciales  $\mathbf{x}(t_0)$ . El vector  $\mathbf{x}(t)$  se denomina *vector de estado del sistema*, mientras la expresión (1.62) corresponde a la *ecuación de estado del sistema*. La señal de salida  $y(t)$  del sistema lineal en análisis se determina por medio del vector de estado en la forma,

$$y(t) = \mathbf{C}\mathbf{x}(t) \quad (1.64)$$

donde la matriz  $\mathbf{C}_{1 \times m}$  se define como,  $\mathbf{C} = [1 \ 0 \ \cdots \ 0]$ .

La generalización de ecuación diferencial (1.59), para los sistemas lineales, corresponde a la ecuación (1.38) e incluye la diferenciación de la señal  $x(t)$ , la cual debe evitarse en el análisis de sistemas estocásticos reales, entre otras razones porque es frecuente en calidad de señal de entrada tomar el ruido blanco, cuyas derivadas en el tiempo no existen.

En este sentido, asumiendo la invariabilidad de los respectivos coeficientes, la ecuación (1.38) se puede representar de forma alterna,

$$\frac{d^m y}{dt^m} + \alpha_{m-1} \frac{d^{m-1} y}{dt^{m-1}} - \beta_{m-1} \frac{du}{dt^{m-1}} + \cdots + \alpha_1 \frac{dy}{dt} - \beta_1 \frac{du}{dt} = \beta_0 u(t) - \alpha_0 y(t), \quad t \geq t_0 \quad (1.65)$$

La integración de ambas partes de la igualdad (1.65), con condiciones iniciales cero, da como resultado,

$$\frac{d^{m-1} y}{dt^{m-1}} + \alpha_{m-1} \frac{d^{m-2} y}{dt^{m-2}} - \beta_{m-2} \frac{d^{m-2} u}{dt^{m-2}} + \cdots + \alpha_1 y(t) - \beta_1 u(t) = x_m(t), \quad t \geq t_0 \quad (1.66)$$

donde

$$x_m(t) = \int_{t_0}^t (-\alpha_0 y(\tau) + \beta_0 u(\tau)) d\tau \quad (1.67)$$

De la misma manera, la ecuación (1.66) se puede escribir en la forma

$$\frac{d^{m-1} y}{dt^{m-1}} + \alpha_{m-1} \frac{d^{m-2} y}{dt^{m-2}} - \beta_{m-1} \frac{d^{m-2} u}{dt^{m-2}} + \cdots + \alpha_2 \frac{dy}{dt} - \beta_2 \frac{du}{dt} = x_m(t) - \alpha_1 y(t) - \beta_1 u(t), \quad t \geq t_0$$

cuya integración conlleva a la ecuación,

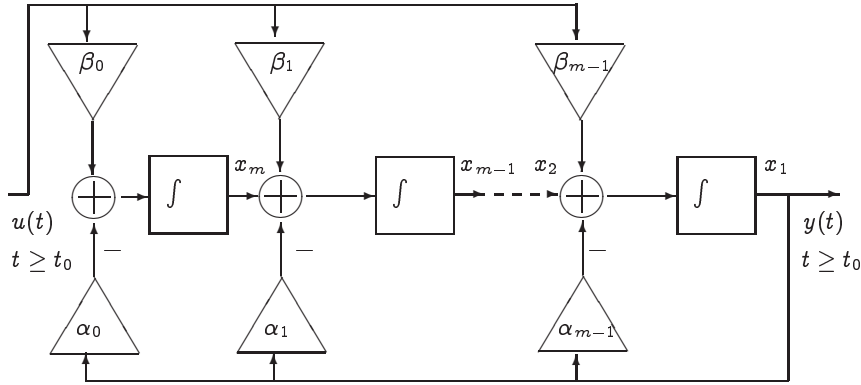
$$\frac{d^{m-2}y}{dt^{m-2}} + \alpha_{m-1} \frac{d^{m-3}y}{dt^{m-3}} - \beta_{m-1} \frac{d^{m-3}u}{dt^{m-3}} + \cdots + \alpha_2 y(t) - \beta_1 u(t) = x_{m-1}(t), \quad t \geq t_0$$

donde

$$x_{m-1}(t) = \int_{t_0}^t (x_m(t) - \alpha_1 y(\tau) + \beta_1 u(\tau)) d\tau \quad (1.68)$$

La sucesión descrita de operaciones, repetida  $m$  veces, da como resultado,

$$y(t) = x_1(t), \quad x_1(t) = \int_{t_0}^t (x_2(t) - \alpha_{m-1} y(\tau) + \beta_{m-1} u(\tau)) d\tau \quad (1.69)$$



**Figura 1.12.** Modelo de un sistema de dimensión múltiple.

El modelo del procesador analógico, que realiza estas transformaciones, se muestra en la Figura (1.12). La salida  $y(t)$ ,  $t \geq t_0$ , se determina si además de tener la entrada  $u(t)$ ,  $t \geq t_0$ , se fijan las condiciones iniciales  $x_k(t_0)$ ,  $k = 1, \dots, m$  de las salidas de los respectivos integradores. Por lo tanto, estas señales se pueden seleccionar en calidad de variables de estado del sistema dinámico descrito por la ecuación (1.38). En correspondencia con las expresiones (1.67), (1.68) y (1.69), las ecuaciones para estas variables de estado son:

$$\begin{aligned} x_1(t) &= y(t) \\ x_2(t) &= \frac{dx_1}{dt} + \alpha_{m-1} y(t) - \beta_{m-1} u(t) \\ &\dots = \dots \\ x_m(t) &= \frac{dx_{m-1}}{dt} + \alpha_1 y(t) - \beta_1 u(t) \\ \frac{dx_m}{dt} &= -\alpha_0 y(t) + \beta_0 u(t) \end{aligned}$$

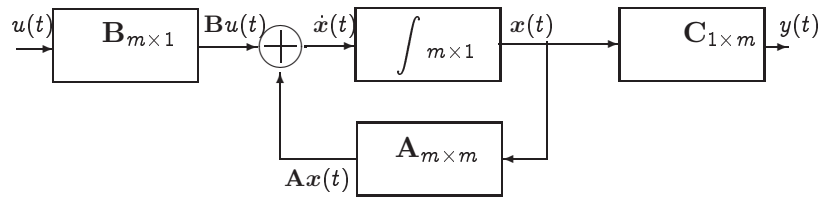


El sistema de ecuaciones de estado obtenidas para el sistema (1.38), al ser representadas mediante la ecuación equivalente de primer orden (1.61), implica de nuevo el cálculo de las correspondientes matrices en (1.63).

En particular, se obtiene que

$$\mathbf{A} = \begin{bmatrix} -\alpha_{m-1} & 1 & 0 & \cdots & 0 \\ -\alpha_{m-2} & 0 & 1 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ -\alpha_1 & 0 & 0 & \cdots & 1 \\ -\alpha_0 & 0 & 0 & \cdots & 0 \end{bmatrix}, \quad \mathbf{B} = \begin{bmatrix} \beta_{m-1} \\ \beta_{m-2} \\ \vdots \\ \beta_1 \\ \beta_0 \end{bmatrix}$$

La representación de la salida sigue siendo la misma que en (1.64). Aunque la ilustración del modelo de sistema de dimensión múltiple de la Figura 1.12, puede ser simplificada, empleando dispositivos matriciales con varias entradas y salidas, como se observa en la Figura 1.13.



**Figura 1.13.** Modelo matricial de un sistema lineal con dimensión  $m$

Las ecuaciones (1.62) y (1.64) se emplean para la representación de sistemas variantes, esto es, con parámetros variables en el tiempo, descritos en forma generalizada por la expresión (1.38). Sin embargo, en este caso las matrices (1.63) son dependientes del tiempo, por lo cual la descripción del modelo, asumiendo  $t \geq t_0$  y dado el vector de condiciones iniciales  $\mathbf{x}(t_0)$ , es:

$$\begin{cases} \frac{d\mathbf{x}}{dt} = \mathbf{A}(t)\mathbf{x}(t) + \mathbf{B}(t)u(t) \\ y(t) = \mathbf{C}(t)\mathbf{x}(t) \end{cases} \quad (1.70)$$

---

**Ejemplo 1.8.** Sea el par de sistemas unidos, como se muestra en la parte superior de la Figura 1.14(a), descritos por el modelo

$$\begin{cases} \frac{d\mathbf{x}_i}{dt} = \mathbf{A}_i(t)\mathbf{x}_i(t) + \mathbf{B}_i(t)u_i(t) \\ y_i(t) = \mathbf{C}_i(t)\mathbf{x}_i(t), \quad \forall t \geq t_0, \quad i \in \{1, 2\}, \quad \mathbf{x}_{i_{r_i \times 1}}(t), \quad r_i \in \{m, n\} \end{cases}$$

Hallar los parámetros del modelo conjunto (parte inferior de la Figura 1.14(a)).

La ecuación vectorial única del sistema conjunto tiene la misma estructura que (1.70) con los

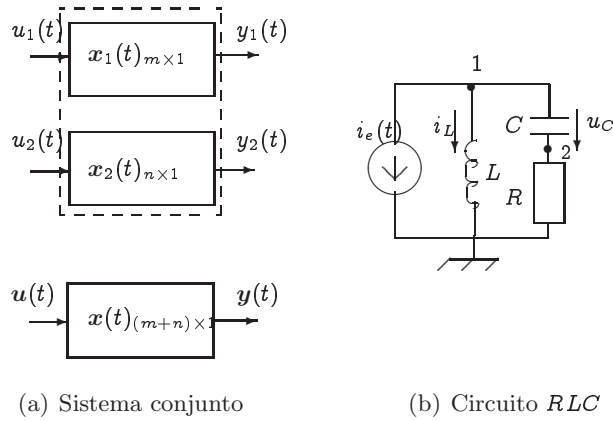


Figura 1.14. Ejemplo de sistemas

siguientes vectores únicos de estado y señales de entrada y salida, respectivamente:

$$\mathbf{x}(t) = \begin{bmatrix} \mathbf{x}_1(t) \\ \vdots \\ \mathbf{x}_2(t) \end{bmatrix}, \quad \mathbf{z}(t) = \begin{bmatrix} \mathbf{z}_1(t) \\ \vdots \\ \mathbf{z}_2(t) \end{bmatrix}, \quad \mathbf{z} \in \{\mathbf{u}, \mathbf{y}\}$$

mientras, la correspondiente estructura de matriz de parámetros tiene la forma,

$$\mathbf{P}(t) = \begin{bmatrix} \mathbf{P}_1(t) & \mathbf{0} \\ \mathbf{0} & \mathbf{P}_2(t) \end{bmatrix}, \quad \mathbf{P} \in \{\mathbf{A}, \mathbf{B}, \mathbf{C}\}, \quad \mathbf{P}_i \in \{\mathbf{A}_i, \mathbf{B}_i, \mathbf{C}_i\}, \quad i \in \{1, 2\}$$

**Ejemplo 1.9.** Hallar las correspondientes matrices del modelo de ecuación de estados del sistema para el circuito mostrado en la Figura 1.14(b).

La ecuación diferencial del circuito  $RLC$  1.14(b) corresponde a la expresión

$$LC \frac{d^2 i_L}{dt^2} + RC \frac{di_L}{dt} + i_L(t) = i_e(t) + RC \frac{di_e}{dt}$$

En calidad de variables de estado se escogen el voltaje en el condensador  $x_1 = u_C$  y la tensión en la bobina,  $x_2 = i_L$ , descritas respectivamente por las siguientes ecuaciones diferenciales:

$$\frac{du_C}{dt} = \frac{1}{C} (i_e(t) - i_L), \quad \frac{di_L}{dt} = \frac{1}{L} (u_C + Ri_e(t) - Ri_L)$$

De las anteriores ecuaciones, se observa que las matrices son constantes en el tiempo, motivo por el cual el modelo (1.70) se puede completar hasta la forma

$$\frac{d\mathbf{x}}{dt} = \mathbf{A}\mathbf{x}(t) + \mathbf{B}\mathbf{u}(t) \tag{1a}$$

$$\mathbf{y}(t) = \mathbf{C}\mathbf{x}(t) + \mathbf{D}\mathbf{u}(t) \tag{1b}$$

luego, para  $\mathbf{u}(t) = i_e(t)$ , se tienen las siguientes matrices:

$$\mathbf{A} = \begin{bmatrix} 0 & -1/C \\ 1/L & -R/L \end{bmatrix}, \quad \mathbf{B} = \begin{bmatrix} 1/C \\ R/L \end{bmatrix}$$

Si en calidad de variables de estado se escogen los voltajes  $y_i \in \{v_1, v_2\}$  sobre los puntos 1 y 2, mostrados en el circuito *RLC* de la Figura 1.14(b), entonces se tienen las ecuaciones:

$$\begin{cases} v_1 &= u_C + v_2 = u_C + R(i_e(t) - i_L), \\ v_2 &= R(i_e(t) - i_L) \end{cases}$$

con lo que las matrices respectivas se describen como  $\mathbf{C} = \begin{bmatrix} 1 & -R \\ 0 & -R \end{bmatrix}$ ,  $\mathbf{D} = \begin{bmatrix} R \\ R \end{bmatrix}$ .

### 1.4.2. Solución de la ecuación de estado

**Ecuación homogénea con coeficientes constantes.** Se asume la ecuación diferencial

$$\frac{d\mathbf{x}}{dt} = \mathbf{A}\mathbf{x}(t), t \geq t_0 \quad (1.72)$$

con condiciones iniciales  $\mathbf{x}(t_0)$ ,  $t \geq t_0$ . Cuando  $\mathbf{x}(t) \Rightarrow \mathbf{x}(t)$ , esto es, se considera una función escalar, para (1.72) se tiene la solución,  $\mathbf{x}(t) = e^{\mathbf{A}(t-t_0)}\mathbf{x}(t_0)$ . La misma estructura de solución se considera para el caso,  $\mathbf{x}(t) = e^{\mathbf{A}(t-t_0)}\mathbf{x}(t_0)$ , considerando la expansión:  $e^{\mathbf{A}t} = \mathbf{I} + \mathbf{A}t + \frac{(\mathbf{A}t)^2}{2!} + \dots + \frac{(\mathbf{A}t)^n}{n!}$ . Además, si la matriz  $\mathbf{A}$  es de orden  $n \times n$ , entonces cualquier polinomio o serie convergente de  $\mathbf{A}$ , puede ser descrita en forma de combinación lineal de los términos  $\mathbf{I}, \mathbf{A}, \dots, \mathbf{A}^n$ .

Sea la notación de la función matricial  $\Phi(t - t_0) \triangleq e^{\mathbf{A}(t-t_0)}$ , denominada *matriz de transición de estado*, y de la cual se exigen las siguientes condiciones:

$$\begin{cases} \Phi(t_0 - t_0) &= \mathbf{I} \\ \frac{d}{dt}\Phi(t - t_0) &= \mathbf{A}(t)\Phi(t - t_0) \\ \frac{d}{dt}\Phi^\top(t, t_0) &= -\mathbf{A}^\top(t_0)\Phi(t, t_0) \end{cases}$$

Por lo tanto, la solución del caso homogéneo tiene forma  $\mathbf{x}(t) = \Phi(t - t_0)\mathbf{x}(t_0)$ .

**Ecuación homogénea con coeficientes variables.** Entonces, las matrices del modelo (1.70) son variantes en el tiempo. Se considera la matriz de transición de estado como función de dos variables,  $\Phi(t, t_0) = e^{\mathbf{A}(t, t_0)}$ , que cumple la siguiente ecuación diferencial:

$$\frac{d}{dt}\Phi(t, t_0) = \mathbf{A}(t)\Phi(t, t_0), \quad \Phi(t_0, t_0) = \mathbf{I}$$

con las siguientes propiedades adicionales, para  $t_0 < t_1 < t_2$ ,

$$\begin{cases} \Phi(t_2, t_0) &= \Phi(t_2, t_1)\Phi(t_1, t_0), \\ \Phi^{-1}(t_1, t_0) &= \Phi(t_0, t_1) \end{cases}$$

En realidad, una expresión compacta para  $\Phi(t, t_0)$  es difícil de obtener. Sin embargo, en la práctica, es más importante saber que ésta expresión existe y que posee ciertas propiedades.

Si es necesario hallar la matriz de transición de estado, se recurre a métodos numéricos. En general, la solución del caso homogéneo con parámetros variables se representa mediante la expresión,  $\mathbf{x}(t) = \Phi(t, t_0)\mathbf{x}(t_0)$ ,  $t \geq t_0$ .

**Ecuación no homogénea con coeficientes variables.** La solución de (1.62) es:

$$\mathbf{x}(t) = \Phi(t, t_0)\mathbf{x}(t_0) + \int_{t_0}^t \Phi(t, \tau)\mathbf{B}(\tau)\mathbf{u}(\tau)d\tau \quad (1.73)$$

donde la matriz de transición de estado  $\Phi(t, t_0)$  cumple las condiciones [11]:

$$\begin{aligned} \Phi(t_0, t_0) &= \mathbf{I}, (\Phi(t, \tau))^{-1} = \Phi(\tau, t), \\ \frac{d}{dt}\Phi(t, \tau) &= \mathbf{A}(t)\Phi(t, \tau) \\ \frac{d}{dt}\Phi^\top(t, \tau) &= -\mathbf{A}^\top(\tau)\Phi^\top(t, \tau) \\ \Phi(t, \tau)\Phi(\tau, \theta) &= \Phi(t, \theta) \end{aligned}$$

La solución (1.73) contiene dos componentes: la primera, que corresponde a las oscilaciones libres, y la segunda, que corresponde a las oscilaciones forzadas. Por cierto, para que el sistema se mantenga estable, el primer término debe atenuarse en el tiempo, de tal manera que el vector de estado se determine, más bien, por la componente forzada, que se obtiene como solución del mismo modelo (1.62), pero con condiciones iniciales 0.

El cálculo de la matriz  $\Phi(t)$  se realiza empleando la transformada de Laplace, (1.25a), sobre (1a), con lo que se obtiene la siguiente ecuación algebraica vectorial,

$$s\mathbf{X}(s) - \mathbf{x}(t_0) = \mathbf{A}^{-1}\mathbf{X}(s) + \mathbf{B}\mathbf{U}(s)$$

Entonces, la imagen del vector de las variables de estado es

$$\mathbf{X}(s) - \mathbf{x}(t_0) = (s\mathbf{I} - \mathbf{A})^{-1}\mathbf{x}(t_0) + (s\mathbf{I} - \mathbf{A})^{-1}\mathbf{B}\mathbf{U}(s)$$

luego,  $\mathbf{x}(t) = \mathcal{L}^{-1}\{(s\mathbf{I} - \mathbf{A})^{-1}\mathbf{x}(t_0)\} + \mathcal{L}^{-1}\{(s\mathbf{I} - \mathbf{A})^{-1}\mathbf{B}\mathbf{U}(s)\}$ .

Asumiendo la invariabilidad de las matrices, entonces la matriz de transición de estado se calcula como,  $\Phi(t) = \mathcal{L}^{-1}\{(s\mathbf{I} - \mathbf{A})^{-1}\}$ .

**Ejemplo 1.10.** Calcular la matriz de transición de estado para un sistema de segundo orden.

Los sistemas lineales se caracterizan por los valores propios  $\lambda_k, k = 1, \dots, n$  de la matriz  $\mathbf{A}$ , que corresponden a las raíces del polinomio característico,  $\det(\lambda\mathbf{I} - \mathbf{A}) = 0$ . Cuando todas las raíces del polinomio son diferentes,  $\lambda_k \neq \lambda_l, \forall k, l = 1, \dots, n$  entonces la matriz de transición de estado es igual a

$$\Phi(t) = e^{\lambda_1 t} \frac{\mathbf{A} - \lambda_2 \mathbf{I}}{\lambda_1 - \lambda_2} - e^{\lambda_2 t} \frac{\mathbf{A} - \lambda_1 \mathbf{I}}{\lambda_1 - \lambda_2}$$

En particular, sea el sistema de segundo orden,  $n = 2$ , descrito por el par de ecuaciones diferenciales  $dx_1/dt = x_2$ ,  $dx_2/dt = -a_1x_1 - a_2x_2 + u(t)$ , con lo que se obtienen las matrices

$$\mathbf{A} = \begin{bmatrix} 0 & 1 \\ -a_1 & -a_2 \end{bmatrix}, \mathbf{B} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

Los valores propios del sistema se determinan de la ecuación,  $\det(\lambda\mathbf{I} - \mathbf{A}) = \lambda^2 + a_2\lambda + a_1 = 0$ , que tiene la solución  $\lambda_{1,2} = -a_2/2 \pm \sqrt{a_2^2/4 - a_1}$ .

El reemplazo de los valores obtenidos para los valores propios en la matriz de transición de estados, en el caso particular de un sistema de segundo orden, da como resultado,

$$\Phi(t) = e^{-a_2t/2} \begin{bmatrix} \frac{a_2}{2\beta} \sin\beta t + \cos\beta t & \frac{1}{\beta} \sin\beta t \\ -\frac{a_1}{\beta} \sin\beta t & -\frac{a_2}{2\beta} \sin\beta t + \cos\beta t \end{bmatrix}, \quad \beta = \sqrt{a_1 - a_2^2/4}$$

Los sistemas lineales variantes en el tiempo con múltiples entradas y salidas, a partir de (1.35), se describen por la matriz cuadrada de la respuesta impulso  $\mathbf{h}(t, \tau)$  en la forma,

$$\mathbf{y}(t) = \int_{-\infty}^t \mathbf{h}(t, \tau) \mathbf{u}(\tau) d\tau \quad (1.74)$$

En muchos casos reales, cuando  $t_0 \rightarrow -\infty$ , en (1.73) se puede despreciar el primer término, con lo cual de (1.62) y (1.73) se obtiene,

$$\mathbf{y}(t) = \mathbf{C}(t) \int_{-\infty}^t \Phi(t, \tau) \mathbf{B}(\tau) \mathbf{u}(\tau) d\tau \quad (1.75)$$

Así mismo, de las expresiones (1.74) y (1.75) se deduce que,

$$\mathbf{h}(t, \tau) = \begin{cases} \mathbf{C}(t) \Phi(t, \tau) \mathbf{B}(\tau), & \tau \leq t \\ \mathbf{0}, & \tau > t \end{cases} \quad (1.76)$$

Cabe anotar que las matrices, que determinan la respuesta a impulso en (1.76), dependen del método elegido de disposición del vector de estado del sistema considerado en cada caso concreto, y por lo tanto, para un mismo sistema las matrices pueden ser diferentes. Sin embargo, la respuesta a impulso matricial  $\mathbf{h}(t, \tau)$  es única para un sistema dado.

En general, existe una transformación lineal de las coordenadas de estado, en la forma  $\mathbf{z} = \mathbf{F}\mathbf{x}$ , tal que del sistema conformado por las ecuaciones (1.62) y (1.64) se obtiene el modelo,

$$\begin{cases} \dot{\mathbf{z}} = \hat{\mathbf{A}}\mathbf{z} + \mathbf{B}\mathbf{u} \\ \mathbf{y} = \hat{\mathbf{C}}^T \mathbf{z} \end{cases}$$

donde

$$\hat{\mathbf{A}} = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & \cdots & 1 \\ -a_n & -a_{n-1} & -a_{n-2} & \cdots & -a_1 \end{bmatrix}, \hat{\mathbf{B}} = \mathbf{F}\mathbf{B}, \hat{\mathbf{C}} = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \mathbf{F} = \begin{bmatrix} \mathbf{C}^\top \\ \mathbf{C}^\top \mathbf{A} \\ \vdots \\ \mathbf{C}^\top \mathbf{A}^{n-1} \end{bmatrix} \quad (1.77)$$

En concordancia con los valores de las matrices en (1.77), el modelo (1.62) se puede reemplazar por la ecuación diferencial escalar del tipo,

$$\frac{d^n y}{dt^n} + a_1 \frac{d^{n-1} y}{dt^{n-1}} + \cdots + a_n y = b_1 \frac{d^{n-1} u}{dt^{n-1}} + \cdots + b_n u, \quad b_i = \sum_{k=0}^{i-1} a_{i-k} \hat{g}_k + \hat{g}_i \quad (1.78)$$

En los sistemas lineales, cuando las matrices  $\mathbf{A}(t)$ ,  $\mathbf{B}(t)$  y  $\mathbf{C}(t)$  muestran dependencia del tiempo, se puede encontrar la correspondiente representación en forma de ecuación diferencial (1.78), con los siguientes coeficientes:

$$b_i(t) = \sum_{k=0}^{i-1} \sum_{l=0}^{i-k} C_{n+l-1}^{n-1} a_{i-k-l} \frac{d}{dt^l} \hat{g}_k(t) + \hat{g}_i(t)$$

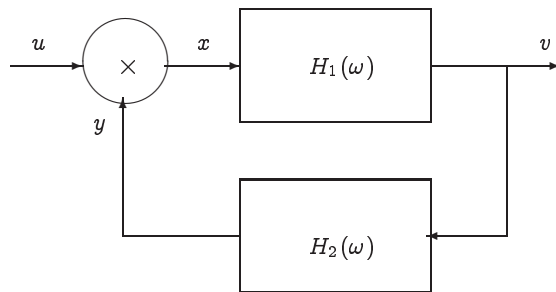
La representación del sistema por las expresiones (1.77) y (1.78), frecuentemente es más cómoda para su análisis mediante procesadores digitales.

## Problemas

**Problema 1.20.** Hallar el modelo equivalente de los circuitos mostrados en la Figura 1.8.

**Problema 1.21.** Hallar la matriz de transición de estados de los circuitos de la Figura 1.8.

**Problema 1.22.** Hallar la estabilidad del sistema con retroalimentación mostrado en la Figura 1.15, con las siguiente funciones de transferencia:  $H_1(\omega) = k/(1 + j\omega)^3$  y  $H_2(\omega) = 1$ .



**Figura 1.15.** Ejemplo de sistema con retroalimentación

# Capítulo 2

## Representación de variables aleatorias

El concepto general de medida  
comprende, como caso particular, el  
concepto de probabilidad

---

Kolmogorov

El desarrollo de la teoría de funciones en espacios métricos implica el estudio de aquellas propiedades de las funciones, las cuales dependen de la medida de conjuntos en los que las funciones toman cualquier grupo de valores. Así por ejemplo, se tienen las propiedades de ortogonalidad de un conjunto de funciones y su completitud, analizadas en §1.1.2. Si se consideran los valores aleatorios como funciones, determinadas en un espacio de eventos elementales, entonces todos métodos de Teoría de Probabilidades aplicadas a estas funciones son un caso específico del empleo de la Teoría Generalizada de Medida. En particular, gran parte de los métodos de análisis probabilístico se basan en el empleo del concepto de independencia de variables aleatorias, que se puede entender como una propiedad de los espacios métricos.

### 2.1. Valores y funciones de probabilidad

La aleatoriedad implica incertidumbre. La representación de las variables aleatorias consiste en la descripción y caracterización de sus principales propiedades que midan esa incertidumbre inherente.

#### 2.1.1. Espacio probabilístico de variables

Cualquier modelo de un experimento analizado a priori se caracteriza por un conjunto de posibles resultados u *observaciones*. El subconjunto formado por un solo elemento se denomina *evento* o *suceso elemental* y todo el conjunto  $\Omega$  se denomina *espacio de eventos elementales*. Los posibles subconjuntos de eventos  $A_k \in \Omega$ , pueden ser de dos clases: eventos

incompatibles, cuando:

$$\bigcup_{i=1}^n A_i = \Omega, \quad A_k \cap A_l = \emptyset, \quad k, l = 1, \dots, n$$

y eventos complementarios,  $\{\bar{A}_k\}$ , tales que

$$A \cup \bar{A} = \Omega \text{ y } A \cap \bar{A} = \emptyset$$

**Definición 2.1.** ( $\sigma$ -álgebra  $\mathcal{B}$ ) El sistema conformado por los subconjuntos de  $\Omega$  se denomina álgebra  $\mathcal{B}$ , si se cumple que para  $\{A_i \in \mathcal{B} : i = 1, \dots, n\}$  se observa la igualdad

$$\bigcup_{i=1}^n A_i = \mathcal{B}$$

esto es, de la sucesión de eventos  $A_i$ , por lo menos ocurre una. Así mismo, se cumple que

$$\bigcap_{i=1}^n A_i = \mathcal{B}$$

por lo que todos los sucesos  $A_k$  ocurren de manera simultánea. La definición se completa con la condición por la cual teniendo  $A \in \mathcal{B}$ , resulta que  $\bar{A} \in \mathcal{B}$ . Por lo anterior, el álgebra  $\mathcal{B}$  es una clase de conjuntos cerrados con relación a una cantidad contable de las operaciones básicas de unión, intersección y complemento ( $\sigma$ -álgebra).

En general, el *espacio probabilístico*  $\{\Omega, \mathcal{B}, P\}$  está conformado por las siguientes tres componentes: el espacio de los eventos elementales  $\Omega$ , la  $\sigma$ -álgebra  $\mathcal{B}$  de los subconjuntos de Borell del espacio  $\Omega$  ( $\sigma$ -álgebra de los eventos) y la estabilidad de las frecuencias de aparición en términos relativos de los eventos o medida de *probabilidad*  $P(A)$ , determinada para  $A \in \mathcal{B}$ , con una masa total unitaria, esto es,  $P(\Omega) = 1$ . La medida de probabilidad se determina a partir de los siguientes principios, denominados *axiomas de Kolmogorov* [12]:

**Axioma I.** A cada evento  $A \in \mathcal{B}$  le corresponde un valor real  $P(A)$ , denominado probabilidad de  $A$ , tal que,  $0 \leq P(A) \leq 1$ .

**Axioma II.**  $P(\Omega) = 1$ .

**Axioma III.** Si  $A \cap B = \emptyset$ , entonces,  $P(A \cup B) = P(A) + P(B)$ .

De los anteriores axiomas se deducen las siguientes propiedades:

- (a). Si  $A$  es un evento imposible, entonces  $P(A) = 0$ .
- (b).  $P(\bar{A}) = 1 - P(A)$ .
- (c). Si  $A \subseteq B$  entonces, se cumple la desigualdad,  $P(A) \leq P(B)$ .
- (d). Si se cumple que  $A_i \cap A_k = \emptyset$ ;  $i, k = 1, \dots, n$ , luego,

$$P\left(\bigcup_{k=1}^n A_k\right) = \sum_{k=1}^n P(A_k) \tag{2.1}$$



En forma general, si  $A_i \cap A_k \neq \emptyset$ ;  $i, k = 1, \dots, n$ , entonces,

$$P\left(\bigcup_{k=1}^n A_k\right) \leq \sum_{k=1}^n P(A_k)$$

(e). Dada cualquier pareja de sucesos  $A$  y  $B$ , se cumple

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

Dos eventos  $A$  y  $B$  son *dependientes* si la probabilidad  $P(A)$  depende de que suceda el evento  $B$ . La probabilidad de la aparición conjunta de ambos eventos se da como

$$P(A \cap B) = P(A)P(B|A) = P(B)P(A|B)$$

Las probabilidades marginales  $P(A)$  y  $P(B)$  se denominan *a priori*, mientras las condicionales  $P(B|A)$  y  $P(A|B)$  - *a posteriori*.

Cuando se asume la independencia mutua de aparición de cada uno de los eventos  $A$  y  $B$ , las respectivas probabilidades a priori y a posteriori se determinan por las expresiones:  $P(B|A) = P(B)$  y  $P(A|B) = P(A)$ , y por lo tanto,  $P(A \cap B) = P(A)P(B)$ , relación que se puede generalizar en la definición de *independencia* del conjunto de sucesos  $\{A_k\}$ :

$$P\left(\bigcap_{k=1}^n A_k\right) = \prod_{k=1}^n P(A_k) \quad (2.2)$$

En cambio, si el conjunto  $\{A_k\}$  es *dependiente* se cumple que:

$$P\left(\bigcap_{k=1}^n A_k\right) = P(A_1)P(A_2|A_1)P(A_3|A_1 \cap A_2) \cdots P\left(A_n \middle| \bigcap_{k=1}^{n-1} A_k\right)$$

**Ejemplo 2.1.** Se reciben 10 números en una serie aleatoria con valores equiprobables del 0 al 9, generados de forma independiente. Hallar la probabilidad de que el primer número obtenido de la serie sea múltiplo de 3.

Los posibles eventos  $A_k$  son 3 (3, 6 y 9), ambos dados con probabilidad de generación igual a  $P(A_k) = 0.1$ , son independientes y excluyentes, por lo tanto, la aparición de cada uno de ellos, acordes con la expresión (2.1), es igual a

$$P(3, 6, 9) = 0.1 + 0.1 + 0.1 = 0.3.$$

**Ejemplo 2.2.** Del anterior ejemplo, hallar la probabilidad de aparición de que por lo menos uno de los valores de la serie aleatoria sea múltiplo de 3.

En este caso, los eventos son independientes pero no excluyentes. La aparición de cualquier suceso  $A_k$  excluye la aparición de la posibilidad de no aparición de todos los demás eventos; situación que se puede analizar como el complemento de (2.2),

$$P\left(\bigcup_{k=1}^n A_k\right) = 1 - P\left(\bigcap_{k=1}^n \bar{A}_k\right) = \prod_{k=1}^n P(\bar{A}_k) = 1 - \prod_{k=1}^n (1 - P(A_k))$$

que en el caso particular de  $n = 10$  y  $P(A_k) = 1/n$ , resulta en  $P(\bigcup_{k=1}^n A_k) = 0.271$ .

La medida de probabilidad  $P(A)$  se completa con los siguientes dos teoremas:

**Teorema 2.2.** (Probabilidad completa) La probabilidad del evento  $A$ , de que aparezca con una de las  $n$  posibles hipótesis  $B_1, \dots, B_n$  o sucesos incompatibles que conforman un grupo topológico completo, esto es,  $A = \bigcup_{k=1}^n (A \cap B_k)$  y teniendo en cuenta que  $(A \cap B_k) \cap (A \cap B_l) = \emptyset, \forall k \neq l$ , se define como

$$\begin{aligned} P(A) &= \sum_{k=1}^n P\left(\bigcup_{k=1}^n (A \cap B_k)\right) = \sum_{k=1}^n P(A \cap B_k) \\ &= \sum_{k=1}^n P(B_k) P(A|B_k) \end{aligned} \quad (2.3)$$

**Teorema 2.3.** (Bayes) Sea el conjunto de hipótesis  $\{B_k : k = 1, \dots, n\}$  un grupo completo, empleando la regla de multiplicación se encuentra que

$$P(B_k|A) = \frac{P(B_k) P(A|B_k)}{P(A)}$$

Si se reemplaza la probabilidad a priori  $P(A)$  por su valor dado en (2.3), se obtiene la siguiente expresión, denominada fórmula de Bayes:

$$P(B_k|A) = \frac{P(B_k) P(A|B_k)}{\sum_{k=1}^n P(B_k) P(A|B_k)} \quad (2.4)$$

**Ejemplo 2.3.** Las resistencias de un sistema se obtienen de dos puntos de producción diferentes: la fábrica  $A_1$  (que produce el 70 %) y la fábrica  $A_2$  (30 %). Se conoce que la fábrica  $A_1$  en promedio produce el 83 % de resistencias con calidad aceptable, mientras la fábrica  $A_2$ , tan solo el 63 %. Determinar las siguientes probabilidades:

1. La probabilidad de que la primera resistencia escogida al azar sea de cada una de las fábricas:  $A_1$  y  $A_2$ .
2. La probabilidad de que las dos primeras resistencias escogidas al azar sean de cada una de las fábricas:  $A_1$  y  $A_2$ .
3. La probabilidad de que por lo menos una resistencia de las dos primeras escogidas al azar sean de cada una de las fábricas:  $A_1$  y  $A_2$ .
4. La probabilidad de que la primera resistencia escogida al azar sea de calidad y producida por cada una de las fábricas:  $A_1$  y  $A_2$ .
5. La probabilidad de que la primera resistencia escogida al azar tenga calidad aceptable.
6. La probabilidad de que si la primera resistencia escogida al azar tiene la calidad aceptable, entonces sea de cada una de las fábricas:  $A_1$  y  $A_2$ .

Los valores respectivos de probabilidad se hallan para cada caso de la siguiente forma,

**caso 1.** Se tiene un solo evento, que por definición es igual a

$$P(A_1) = \frac{70}{100} = 0.7, \quad P(A_2) = 0.3$$

**caso 2.** Se tienen dos eventos independientes y no excluyentes. La probabilidad de que ambos ocurran se calcula por la expresión (2.2),

$$\begin{aligned} P(A_1(1) \cap A_1(2)) &= 0.7 \cdot 0.7 = 0.49 \\ P(A_2(1) \cap A_2(2)) &= 0.3 \cdot 0.3 = 0.09 \end{aligned}$$

**caso 3.** La probabilidad de que se tenga por lo menos una resistencia de dos hechas por una de las fábricas,  $A_1$  y  $A_2$ , se calcula por la expresión obtenida en el ejemplo 2.2, respectivamente como:

$$P(A_1(1) \cup A_1(2)) = 1 - (1 - P(A_1))(1 - P(A_1)) = 1 - (1 - 0.7)^2 = 0.91$$

$$P(A_2(1) \cup A_2(2)) = 1 - (1 - 0.3)^2 = 0.09$$

**caso 4.** Los eventos considerados son dependientes y no excluyentes; la resistencia de buena calidad  $c$  hecha por cada una de las fábricas  $A_i$ ,  $i = 1, 2$ , con probabilidad condicional  $P_{A_i}(c)$ , cuyo valor de probabilidad se halla respectivamente por la expresión:

$$P(A_1 \cup c) = P(A_1) P_{A_1}(c) = 0.7 \cdot 0.83 = 0.581$$

De igual manera se tiene,  $P(A_2 \cup c) = P(A_2) P_{A_2}(c) = 0.3 \cdot 0.63 = 0.189$ .

**caso 5.** En el caso de selección de una resistencia con aceptable calidad  $c$ , como hipótesis  $B_k$  se asume que la resistencia se hizo en cada una de las fábricas:  $A_i$  ( $P(B_k) = P(A_k)$ , para  $k = 1, 2$ ). Empleando la expresión de probabilidad completa (2.3) se tiene,

$$\begin{aligned} P(c) &= P(A_1 \cap c \cup A_2 \cap c) = P(B_1) P_{A_1}(c) + P(B_2) P_{A_2}(c) \\ &= 0.7 \cdot 0.83 + 0.3 \cdot 0.63 = 0.77 \end{aligned}$$

**caso 6.** El valor de la probabilidad de la hipótesis  $B_k$  (la resistencia se hizo en la fábrica  $A_k$ ) con la condición de que la resistencia se termina con calidad aceptable se calcula por el teorema de Bayes (2.4),

$$\begin{aligned} P_c(A_1) &= \frac{P(A_1) P_{A_1}(c)}{P(A_1) P_{A_1}(c) + P(A_2) P_{A_2}(c)} = \frac{0.7 \cdot 0.83}{0.7 \cdot 0.83 + 0.3 \cdot 0.63} = 0.755 \\ P_c(A_2) &= \frac{P(A_2) P_{A_2}(c)}{P(A_1) P_{A_1}(c) + P(A_2) P_{A_2}(c)} = \frac{0.3 \cdot 0.63}{0.7 \cdot 0.83 + 0.3 \cdot 0.63} = 0.245 \end{aligned}$$

Cuando se establece alguna regla o relación funcional que asigna un número real a cada posible observación o trayectoria de la *variable aleatoria asignada*, la totalidad de sus valores posibles constituye el conjunto  $\{x_i \in \mathbb{R} : i = 1, 2, \dots\}$  confinado en un espacio  $\mathfrak{X} \in \mathbb{R}$  y conocido como espacio de observaciones o *muestra*. Las variables aleatorias pueden ser discretas o continuas dependiendo de la naturaleza de las observaciones del experimento. Si en cualquier intervalo finito del eje real, la variable  $\xi$  puede asumir solamente un número finito de valores distintos y a priori conocidos, se conoce a la variable como *aleatoria discreta*. Si, por el contrario,  $\xi$  puede tomar cualquier valor en un intervalo dado del eje real, se denomina *variable aleatoria continua*. En este caso, en cualquier intervalo del eje real la cantidad de valores que puede tomar  $\xi$  es infinita.

Un valor aleatorio en  $\{\Omega, \mathcal{B}, P\}$  corresponde a cualquier  $\mathcal{B}$ -función medible  $\xi(\omega) \in \Omega$  (se nota simplemente por  $\xi$ ). La distribución de la variable aleatoria  $\xi$  corresponde a la medida  $F_\xi(x) = P_\xi(\xi < x : \xi \in \mathcal{B})$ . Cuando se cumple que el valor de probabilidad

$$P(\xi = x) = F_\xi(x + \Delta x) - F_\xi(x) > 0, \quad \Delta x \rightarrow 0,$$

entonces  $x$  se denomina el *átomo* de la distribución.

La variable  $\xi$  tiene distribución continua y absoluta si existe la función  $p_\xi > 0$ , que

sea normalizada en el sentido en que  $\int_{\Lambda} p_{\xi}(\lambda) d\lambda = 1$ , denominada *función densidad de probabilidad* (FDP), para la cual se cumple que:

$$\begin{cases} F_{\xi}(x) &= \int_{-\infty}^x p_{\xi}(\lambda) d\lambda, \\ p_{\xi}(\lambda) &= dF_{\xi}(\lambda)/d\lambda \end{cases} \quad (2.5)$$

Se considera que una variable aleatoria  $\xi_n$ ,  $n = 1, 2, \dots$ , converge en el sentido probabilístico a un valor o función  $\xi$ , si para cualquier valor tan pequeño como sea de  $\varepsilon > 0$ , la probabilidad de ocurrencia de la desigualdad corresponde al límite

$$\lim_{\substack{n \rightarrow \infty \\ \varepsilon \rightarrow 0}} P(|\xi_n - \xi| < \varepsilon) = 1, \quad \forall \varepsilon > 0$$

### 2.1.2. Valores medios y momentos

Una forma adecuada para describir una variable aleatoria  $\xi$  corresponde a los valores y momentos que determinan las particularidades de su respectiva FDP,  $p_{\xi}(x)$ .

Los *momentos iniciales* se definen por la siguiente expresión:

$$E\{\xi^n\} \triangleq \int_{-\infty}^{\infty} \xi^n dF_{\xi}(x) = \int_{-\infty}^{\infty} x^n p_{\xi}(x) dx = m_{n\xi}, \quad n \in \mathbb{N} \quad (2.6)$$

**Momento de primer orden.** Valor medio conocido como la *esperanza matemática*:

$$m_{1\xi} = E\{\xi\} = \int_{-\infty}^{\infty} x p_{\xi}(x) dx \quad (2.7)$$

Si la variable aleatoria es discreta, la integral (2.7) se reduce a la sumatoria en la forma:  $m_{1\xi} = \sum_i^N x_i P(\xi = x_i)$ . Por cierto, cuando los eventos son equiprobables,  $x_i = 1/N, \forall i$ , entonces, el primer momento inicial corresponde al *promedio*.

Las siguientes son las principales propiedades del valor medio

$$\begin{cases} E\{c\} &= c, \quad c = \text{const} \\ E\{c\xi\} &= cm_{1\xi}, \quad c = \text{const} \\ E\{\xi \pm \eta\} &= m_{1\xi} \pm m_{1\eta} \end{cases}$$

El concepto de esperanza matemática de la variable aleatoria  $\xi$  se puede generalizar, para el caso de cualquier función determinística  $f(\xi)$ , en la forma:

$$E\{f(\xi)\} = \int_{-\infty}^{\infty} f(x) dF_{\xi}(x) = \int_{-\infty}^{\infty} f(x) p_{\xi}(x) dx \quad (2.8)$$

De otra parte, a fin de evitar la influencia que tiene el valor medio en el cálculo de los momentos de orden superior  $n > 1$  se definen los *momentos centralizados* de orden  $n$ :

$$E\{(\xi - m_{1\xi})^n\} = \int_{-\infty}^{\infty} (x - m_{1\xi})^n p_{\xi}(x) dx = \mu_{n\xi}, \quad \forall n \geq 2, n \in \mathbb{N} \quad (2.9)$$

**Momentos de segundo orden.** Corresponde al *valor cuadrático medio*,

$$m_{2\xi} = \int_{-\infty}^{\infty} x^2 p_{\xi}(x) dx, \quad (2.10)$$

De forma análoga, la sumatoria  $m_{2\xi} = \sum_i^N x_i^2 P_{\xi}(\xi = x_i)$  determina este momento para las variables discretas.

El momento central (2.9) para  $n = 2$  se conoce como *varianza*:

$$\begin{aligned} \mu_{2\xi} &= \int_{-\infty}^{\infty} (x - m_{1\xi})^2 p_{\xi}(x) dx = \int_{-\infty}^{\infty} (x^2 - 2xm_{1\xi} + m_{1\xi}^2) p_{\xi}(x) dx \\ &= \int_{-\infty}^{\infty} (x^2) p_{\xi}(x) dx - 2m_{1\xi} \int_{-\infty}^{\infty} (x) p_{\xi}(x) dx + m_{1\xi}^2 \int_{-\infty}^{\infty} p_{\xi}(x) dx \end{aligned}$$

Teniendo en cuenta las definiciones, (2.7) y (2.10), se obtiene

$$\mu_{2\xi} = m_{2\xi} - 2m_{1\xi}^2 + m_{1\xi}^2 = m_{2\xi} - m_{1\xi}^2 \triangleq \sigma_{\xi}^2 \quad (2.11)$$

Si la magnitud aleatoria es discreta, cuando ésta toma los valores  $\xi_k$  con las probabilidades asociadas  $P(\xi_k)$ , a partir de (2.11), la respectiva varianza se calcula como:

$$\sigma_{\xi}^2 = \sum_{k=1}^N x_k^2 P_{\xi}(\xi = x_k) - \left( \sum_{k=1}^N x_k P_{\xi}(\xi = x_k) \right)^2$$

El valor  $\sigma_{\xi}^2$  caracteriza el nivel de concentración de la FDP,  $p_{\xi}(x)$ , en los alrededores de la esperanza matemática; interpretación que resulta de la *desigualdad de Chebyshev*:

$$P_{\xi}(|x - m_{1\xi}| \geq \varepsilon) \leq \sigma_{\xi}^2 / \varepsilon^2 \quad (2.12)$$

esto es, al disminuir el nivel de dispersión de la variable aleatoria  $\xi$ , aumenta la probabilidad de que sus valores no salgan fuera de los límites del intervalo  $[m_{1\xi} - \varepsilon, m_{1\xi} + \varepsilon]$ , siendo

$\varepsilon \geq 0$ . Cabe anotar que si la FDP es simétrica,  $p_\xi(x) = p_\xi(-x)$ , entonces,  $E\{\xi\} = 0$ , luego,

$$\begin{aligned}\sigma_\xi^2 &= \int_{-\infty}^{\infty} (x - m_{1\xi})^2 p_\xi(x) dx = \int_{-\infty}^{\infty} x^2 p_\xi(x) dx \\ &= 2 \int_0^{\infty} x^2 p_\xi(x) dx\end{aligned}\quad (2.13)$$

En general, la varianza tiene las siguientes propiedades:

$$\begin{cases} \sigma_\xi^2 \geq 0, \text{ si } \sigma_\xi^2 = 0 \Rightarrow \xi \rightarrow \text{const.} \\ \sigma_{c\xi}^2 = \int_{-\infty}^{\infty} (cx - cm_{1\xi})^2 p_\xi(x) dx = c^2 \sigma_\xi^2, \text{ por cierto, } \sigma_{\xi+c}^2 = \sigma_\xi^2, \forall c = \text{const.} \\ \sigma_{\xi\eta}^2 = E\{\xi^2 \eta^2\} = E\{\xi^2\} E\{\eta^2\} = \sigma_\xi^2 \sigma_\eta^2 \end{cases} \quad (2.14)$$

**Ejemplo 2.4.** Una moneda se lanza tres veces. Determinar la función acumulativa correspondiente a todos los posibles sucesos.

De forma condicional, cuando la moneda muestre la cara se denota por 0 y cuando muestre el sello, por 1. Asegurando la independencia de los lanzamientos, los sucesos se agrupan de acuerdo a la cantidad de caras y sellos que muestren sin importar su orden de salida: suceso cero: que haya 3 caras y ningún sello; ocurriría una vez de ocho posibles (1/8); suceso uno: que haya 2 caras y un sello; ocurriría tres veces de ocho posibles (3/8); suceso dos: que haya 1 cara y 2 sellos; saldría tres veces de ocho posibles (3/8); suceso tres: que no haya ninguna cara y tres sellos: una vez de ocho posibles (1/8). De acuerdo al orden descrito de los sucesos, la variable aleatoria toma los valores 0, 1, 2, 3, con función distributiva (representada en la parte inferior de la Figura 2.1(a)) descrita como:

$$F_\xi(x) = (u(x) + 3u(x-1) + 3u(x-2) + u(x-3))/8$$

La FDP respectiva está representada en la parte superior de la Figura 2.1(a) que corresponde a la derivada de la función distributiva e igual a:

$$p_\xi(x) = (\delta(x) + 3\delta(x-1) + 3\delta(x-2) + \delta(x-3))/8$$

**Ejemplo 2.5.** Dada la señal binaria periódica aleatoria, representada en la parte superior de la Figura 2.1(b). Hallar los 2 primeros momentos iniciales y la varianza.

Por cuanto,  $p_\xi(x) = P_1\delta(x-x_1) + P_2\delta(x-x_2)$ , donde  $P_1 = 1 - P_2$ , entonces, los respectivos momentos se calculan como:

$$\begin{aligned}m_{1\xi} &= \int_{-\infty}^{\infty} x (P_1\delta(x-x_1) + P_2\delta(x-x_2)) dx = P_1x_1 + P_2x_2; \\ m_{2\xi} &= \int_{-\infty}^{\infty} x^2 p_\xi(x) dx = \int_{-\infty}^{\infty} x^2 (P_1\delta(x-x_1) + P_2\delta(x-x_2)) dx = P_1x_1^2 + P_2x_2^2; \\ \sigma_\xi^2 &= P_1P_2(x_1-x_2)^2 = P_1(1-P_1)(x_1-x_2)^2\end{aligned}$$

Cuando  $P_1 = P_2 = 0.5 \Rightarrow \sigma_\xi^2 = (x_1 - x_2)^2/4$ , asumiendo  $\xi_1 = -\xi_2 = 1 \Rightarrow \sigma^2 = 1$ .

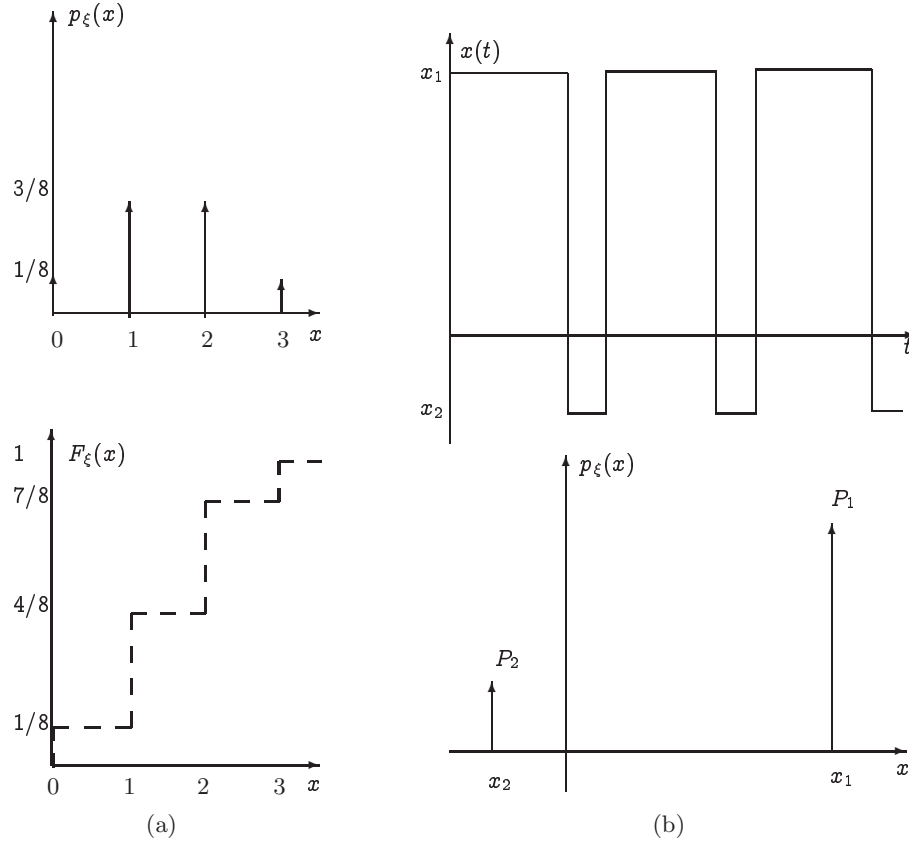


Figura 2.1. Ejemplo de FDP y función de distribución

**Definición 2.4.** Un conjunto de valores aleatorios  $\xi \in \mathbb{C}$ , determinados en un espacio probabilístico  $\{\Omega, \mathcal{B}, P\}$ , tal que su segundo momento sea finito,  $E\{\xi^2\} < \infty$ , conforma un espacio lineal normalizado de Hilbert  $L^2\{\Omega, \mathcal{B}, P\}$  provisto del producto escalar,

$$\langle \xi, \eta \rangle = E\{\xi \eta^*\}$$

con norma  $\|\xi\|_2 = (E\{\xi^2\})^{1/2}$ , mediante la cual se determina la distancia entre las respectivas variables aleatorias:

$$d(\xi, \eta) = \|\xi - \eta\|_2, \quad \xi, \eta \in L^2\{\Omega, \mathcal{B}, P\}$$

Además, toda variable aleatoria  $\xi$ , perteneciente a  $L^2\{\Omega, \mathcal{B}, P\}$ , es una variable aleatoria de Hilbert.

**Definición 2.5.** Las variables aleatorias de Hilbert,  $\xi$  y  $\eta$ , se definen ortogonales cuando se cumple que  $E\{\xi \eta^*\} = 0$ .

**Momento de tercer orden.** El momento central de tercer orden  $\mu_{3\xi}$  puede servir de criterio para estimar el grado de asimetría de la FDP con respecto al eje de abscisa que pasa por la coordenada con valor igual al primer momento de la variable aleatoria.

Frecuentemente, se emplea el *coeficiente de asimetría* definido por la siguiente relación de momentos de tercer orden:

$$\begin{aligned}\gamma_{1\xi} &= \frac{\mu_{3\xi}}{\sqrt{\mu_{2\xi}^3}} = \frac{\mu_{3\xi}}{\sigma_\xi^3} \\ &= \frac{m_{3\xi} - 3m_{1\xi}m_{2\xi} + 2m_{1\xi}^3}{\sigma_\xi^3}\end{aligned}\quad (2.15)$$

Cabe anotar, que cuando la FDP es simétrica, entonces  $\mu_{3\xi} = 0$  y, por lo tanto,  $\gamma_{1\xi} = 0$ . Por lo tanto, se puede juzgar que entre mayor sea el valor del coeficiente, mayor es la asimetría de la FDP.

**Momento de cuarto orden.** Se determina el *coeficiente de exceso*,

$$\begin{aligned}\gamma_{2\xi} &= \frac{\mu_{4\xi}}{\mu_{2\xi}^2} - 3 \\ &= \frac{m_{4\xi} - 4m_{2\xi}m_{1\xi} + 6m_{2\xi}m_{1\xi}^2 - 3m_{1\xi}^4}{\sigma_\xi^4} - 3,\end{aligned}\quad (2.16)$$

el cual es igual a 0 cuando corresponde a la FDP Gaussiana. El exceso compara el comportamiento del pico principal de cualquier FDP simétrica ( $\gamma_{1\xi} = 0$ ) con el pico de la densidad Gaussiana. Si resulta que  $\gamma_{2\xi} > 0$ , entonces el pico principal de la FDP en análisis es más grande y agudo que el pico de la densidad Gaussiana; asumiendo ambas FDP con igual valor medio y varianza.

El momento de orden  $n$  para la variable aleatoria  $|\xi|$  es su *momento absoluto de orden  $n$* , aunque los momentos de mayor orden se emplean con muy poca frecuencia. Sin embargo, es importante tener en cuenta que cualquiera que sea el orden del momento, su valor es único (en caso de existir) para una FDP dada. La afirmación inversa, según la cual a un conjunto dado de momentos la FDP se determina de forma única, no siempre es cierta (*problema de los momentos* [13]). Aunque en la mayoría de los casos prácticos, la afirmación se cumple en ambos sentidos.

Por último, cabe anotar que para algunas FDP, los momentos definidos tanto en (2.6) como en (2.9), puede ocurrir que no existan para ciertos valores de  $n$ .

**Cumulantes.** Basados en la definición dada, en la ecuación (2.7), se determina el valor medio de la función  $e^{j\omega\xi}$ , el cual se denomina *función característica*:

$$\Theta(j\omega) \triangleq E\{e^{j\omega\xi}\} = \int_{-\infty}^{\infty} p_\xi(x) e^{j\omega x} dx \quad (2.17)$$

que en esencia corresponde a la función de densidad espectral (1.18) de la función densidad de probabilidad, aunque con el signo invertido en el exponente.



Mediante la función característica también se pueden determinar los valores de los momentos iniciales, en particular, al diferenciar para  $\omega = 0$ ,

$$\begin{aligned} \left. \frac{d^k \Theta}{d\omega^k} \right|_{\omega=0} &= j^k \int_{-\infty}^{\infty} x^k e^{j\omega x} p_{\xi}(x) dx \bigg|_{\omega=0} = j^k \int_{-\infty}^{\infty} x^k p_{\xi}(x) dx \\ &= j^k m_{k\xi} \end{aligned}$$

de lo cual resulta que,

$$m_{k\xi} = \frac{1}{j^k} \left. \frac{d^k \Theta}{d\omega^k} \right|_{\omega=0}$$

Luego, los momentos (2.6) se pueden definir mediante el valor medio de la descomposición por la serie de MacLaurin:

$$\Theta(j\omega) = E \left\{ \sum_{k=0}^{\infty} \frac{1}{k!} \xi^k (j\omega)^k \right\} = 1 + \sum_{k=1}^{\infty} \frac{m_{k\xi}}{k!} (j\omega)^k$$

Los valores *semi-invariantes* o *cumulantes*  $m'_{k\xi}$  se definen del anterior promedio al evitar la influencia de la constante, para lo cual en (2.17) se halla el promedio de la función transformada en la forma  $\Psi(j\omega) = \ln \Theta(j\omega)$ :

$$\begin{cases} m'_{k\xi} &= j^{-k} \left( \frac{d^k \Psi(j\omega)}{d\omega^k} \right) \bigg|_{\omega=0} \\ \Psi(j\omega) &= \sum_{k=1}^{\infty} \frac{m'_{k\xi}}{k!} (j\omega)^k \end{cases} \quad (2.18)$$

Se observa de la expresión (2.18) que

$$\dot{\Psi}(j\omega) = \frac{\dot{\Theta}(j\omega)}{\Theta(j\omega)}, \quad \ddot{\Psi}(j\omega) = \frac{\Theta(j\omega)\ddot{\Theta}(j\omega) - (\dot{\Theta}(j\omega))^2}{\Theta^2(j\omega)}$$

De la definición de la función característica, se tiene que  $\Theta(0) = 1$ , luego,

$$\dot{\Psi}(0) = \dot{\Theta}(0), \quad \ddot{\Psi}(0) = \ddot{\Theta}(0) - (\dot{\Theta}(0))^2$$

Teniendo en cuenta la expresión (2.17), se obtienen los primeros dos momentos,

$$\begin{cases} \dot{\Theta}(0) &= jm_{1\xi}, \\ \ddot{\Theta}(0) &= -m_{2\xi}, \\ \ddot{\Psi}(0) &= -m_{2\xi} - (jm_{1\xi})^2 = -(m_{2\xi} - m_{1\xi}^2) = -\mu_{2\xi} \\ &= -\sigma_{\xi}^2 \end{cases} \quad (2.19)$$

Los cumulantes y momentos iniciales se relacionan mediante las siguientes expresiones:

$$\begin{aligned} m'_{1\xi} &= m_{1\xi}, m'_{2\xi} = m_{2\xi} - m_{1\xi}^2 = \sigma_\xi^2, m'_{3\xi} = m_{3\xi} - 3m_{1\xi}m_{2\xi} + 2m_{1\xi}^3, \dots \\ m_{1\xi} &= m'_{1\xi}, m_{2\xi} = m'_{2\xi} + m_{1\xi}^2, m_{3\xi} = m'_{3\xi} + 3m'_{1\xi}m'_{2\xi} + m_{1\xi}^3, \dots \end{aligned}$$

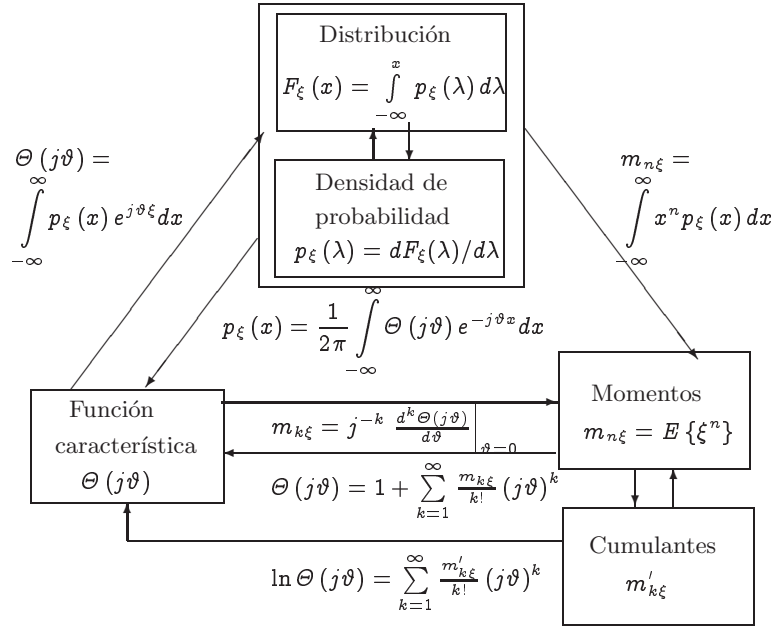
La asimetría y exceso se determinan a través de los cumulantes, respectivamente como,

$$\begin{cases} \gamma_{1\xi} &= m'_{3\xi} m'_{2\xi}^{-3/2}, \\ \gamma_{2\xi} &= m'_{4\xi} m'_{2\xi}^{-2} \end{cases} \quad (2.20)$$

Entre otros valores medios importantes están la *moda* ( $mo_\xi$ ), definida como el valor con mayor probabilidad de aparición de una variable aleatoria  $\xi$ , al cual le corresponde el máximo de la FDP y la *mediana* ( $me_\xi$ ), definida como el valor que satisface la condición:

$$P(\xi < me_\xi) = \int_{-\infty}^{me_\xi} p_\xi(x) dx = \int_{me_\xi}^{\infty} p_\xi(x) dx = \frac{1}{2}$$

Por cierto, si la FDP es una función simétrica, entonces,  $\gamma_{1\xi} = 0$ , además, se cumple que  $\text{sgn}(\gamma_{1\xi}) = \text{sgn}(me_\xi - mo_\xi)$ ,  $\forall \xi \neq 0$ .



**Figura 2.2.** Relación entre la distribución y las características de aleatoriedad

Las diversas relaciones de cálculo que existen entre la función de distribución (o la FDP), la función característica, los momentos y los cumulantes, se muestran en la Figura 2.2.

### 2.1.3. Funciones de probabilidad

Como se antes se dijo, en la conformación del espacio probabilístico de señales, las variables aleatorias, que tienen regularidad estadística, se pueden describir por medio de las funciones de probabilidad definidas en términos de  $\xi$ , en particular, las siguientes:

**Función de distribución acumulativa,  $F_\xi(x)$ .** Se define como la probabilidad de que la variable aleatoria continua tome valores menores que  $\xi = x_0$ :

$$F_\xi(x) = P\{\infty < \xi \leq x\} = P\{\xi \leq x\}$$

Mientras en el caso de variables aleatorias discreta  $\{x_i\} \in \xi$  con las respectivas probabilidades asociadas,  $P_i = P(\xi = x_i)$ , la distribución se expresa como:

$$F_\xi(x) = \sum_{i=1}^N P_i u(\xi - \xi_i)$$

La función de distribución acumulativa tiene las siguientes propiedades:

- (a).  $0 \leq F_\xi(x) \leq 1, \forall x$
- (b).  $F_\xi(x_m) \leq F_\xi(x_n), \forall x_m < x_n,$
- (c).  $F_\xi(-\infty) = 0$ , mientras  $F_\xi(\infty) = 1$ .

**Función densidad de probabilidad.** La variable aleatoria  $\xi$ , que tiene función de distribución absolutamente continua, esto es, la probabilidad de ocurrencia de cualquiera de su cantidad infinita de valores, sobre un intervalo tan angosto como se quiera, es infinitamente pequeña, entonces, existe una función  $p_\xi(x)$ , tal que

$$F_\xi(x) = \int_{-\infty}^x p_\xi(x) dx, \quad (2.21)$$

La función  $p_\xi(x)$  que cumple con (2.21) es la función densidad de probabilidad de la variable continua  $\xi$ , y corresponde al límite de la relación de la probabilidad de ocurrencia de los valores de la variable, dados en el intervalo  $[x, x + \Delta x]$ , sobre la longitud del mismo, cuando  $\Delta x \rightarrow 0$ :

$$p_\xi(x) = \lim_{\Delta x \rightarrow 0} P\{x \leq \xi < x + \Delta x\} / \Delta x$$

de lo anterior resulta que

$$P\{x \leq \xi < x + \Delta x\} = p_\xi(x) \Delta x + \varepsilon\{x\}, \quad \varepsilon\{x\} \geq 0 \quad (2.22)$$

siendo  $\varepsilon\{x\}$  una cantidad residual de mucho menor valor que  $\Delta x$ . La cantidad  $p(x) \Delta x$  corresponde al átomo de la FDP.

Las principales propiedades de la FDP son las siguientes:

- (a).  $p_{\xi}(x) \geq 0, \forall x \in \xi$
- (b).  $\int_{\xi} p_{\xi}(x) dx = F_{\xi}(\max \{x\}) = 1,$
- (c).  $\int_{x_1}^{x_2} p_{\xi}(x) dx = F_{\xi}(x_2) - F_{\xi}(x_1) = P\{x_1 \leq \xi < x_2\}$

Cabe anotar, que en el caso de que no exista la derivada (2.5), para la definición de  $p_{\xi}(x)$ , se recurre a la operación de límite.

De otra parte, para las variables aleatorias discretas se tiene que:

$$p_{\xi}(\xi) = \frac{d}{dx} \left( \sum_i P_i u(x - x_i) \right) = \sum_i P_i \delta(x - x_i),$$

La FDP puede ser discreto-análoga cuando se determina de forma combinada: sobre una malla de valores  $\{x_i \in \xi : i = 1, \dots, n\}$  con las respectivas probabilidades  $\{P_i : i = 1, \dots, n\}$  y sobre un intervalo  $\bar{x} \in \xi$ , para el cual se define un conjunto continuo de valores con densidad de probabilidad  $p_{\bar{x}}(x)$ , con lo cual la FDP de la variable discreto-análoga es la suma ponderada (donde los coeficientes de peso  $a, b$  deben cumplir la condición  $a + b = 1$ ):

$$p_{\xi}(x) = ap_{\bar{x}}(x) + b \sum_{i=1}^n P_i \delta(x - x_i)$$

**Ejemplo 2.6.** Un punto realiza la trayectoria oscilatoria  $x = a \sin \omega t$ . Hallar la FDP de la variable  $x$  en cualquier momento  $t$ , asumiendo, que la probabilidad de encontrarse el punto en el intervalo  $(x, x + dx)$  es proporcional a la longitud del intervalo  $dx$  e inversamente proporcional a la velocidad en el correspondiente momento de tiempo. Hallar la distribución integral y la probabilidad de que el punto se encuentra en el intervalo  $-a + b \leq x \leq a - c$ , ( $b < 2a; c < 2a$ ).

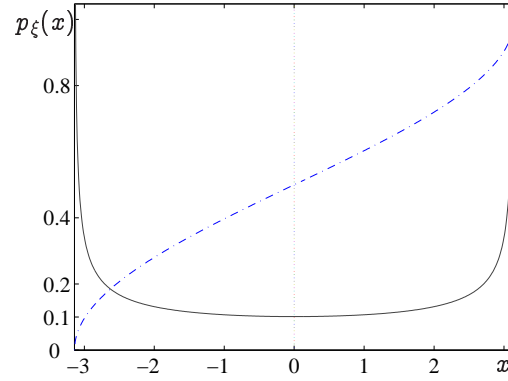
La condición de proporcionalidad para la probabilidad de encontrarse el punto en un intervalo dado se interpreta como,

$$P\{x \leq \xi \leq x + dx\} = k \frac{dx}{dx/dt} = k dt$$

por lo cual,  $p_{\xi}(x) = \frac{d}{dx} P\{\xi\} = k \frac{dt}{dx}$ . Por cuanto,  $dx/dt = a\omega \cos \omega t$ , entonces,

$$\frac{dt}{dx} = \frac{1}{a\omega \cos \omega t} = \frac{1}{a\omega \sqrt{1 - \sin^2 \omega t}} = \frac{1}{a\omega \sqrt{1 - \left(\frac{x}{a}\right)^2}} = \frac{1}{\omega \sqrt{a^2 - x^2}}$$

Asumiendo que  $|x| < a$ , la FDP se determina como,  $p_{\xi}(x) = k/(\omega \sqrt{a^2 - x^2})$ , donde la



**Figura 2.3.** Cálculo de la FDP — y distribución integral —.

constante  $k$  se calcula de la condición de unidad de área para cualquier FDP,

$$\int_{-a}^a \frac{k}{\omega \sqrt{a^2 - x^2}} dx = \frac{k}{\omega} \arccos \frac{x}{a} \Big|_{-a}^a = \frac{k}{\omega} \pi = 1$$

entonces,  $k = \omega/\pi$ , con lo que finalmente se obtiene que,  $p_{\xi}(x) = (\pi \sqrt{a^2 - x^2})^{-1}$ .

La Figura 2.3 muestra la función densidad de probabilidad obtenida (línea continua), en la cual se observa que su valor mínimo es  $\min\{p_{\xi}(x)\} = 1/\pi a$  que ocurre para  $x = 0$ , cuando la velocidad del punto es la máxima posible, mientras en los valores de  $x = \pm a$ , la densidad tiende a infinito, entonces, en esos puntos la FDP no tiene sentido.

La distribución integral (línea punteada), que existe en todo el intervalo de análisis, se calcula de la siguiente integral:

$$\begin{aligned} F_{\xi}(x) &= \int_{-a}^x p(x) dx = \int_{-a}^x \frac{dx}{\pi \sqrt{a^2 - x^2}} = \frac{1}{\pi} \arcsin \frac{x}{a} \Big|_{-a}^x \\ &= \frac{1}{\pi} \left( \arcsin \frac{x}{a} - \left( -\frac{\pi}{2} \right) \right) = \frac{1}{2} + \arcsin \frac{x}{a} \end{aligned}$$

**Variables aleatorias Gaussianas.** En la práctica, el modelo de FDP más empleado corresponde a la densidad *Normal* o *Gaussiana* (Figura 2.4(a)):

$$p(\xi) = \frac{1}{\sqrt{2\pi b}} \exp \left( -\frac{(\xi - a)^2}{2b^2} \right) \triangleq \mathcal{N}(a, b), \quad -\infty < \xi < \infty, \quad (2.23)$$

Los primeros momentos de la FDP Gaussiana, en concordancia con el modelo (2.23), se obtienen empleando la respectiva función característica,  $\Theta(j\omega) = \exp(j\omega a - b^2\omega^2/2)$ , mientras la función logarítmica característica es igual a,  $\Psi(j\omega) = j\omega a - b^2\omega^2/2$ .

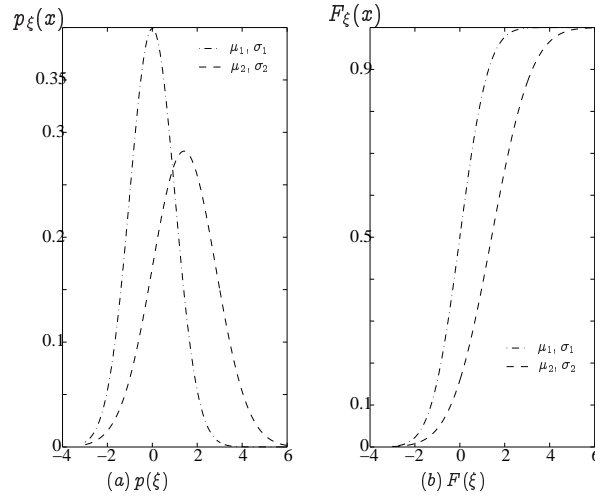
Por lo tanto, el momento inicial de primer orden, de acuerdo con (2.19), es igual a

$$m_{1\xi} = j^{-1} \dot{\Theta}(0) = j^{-1} \dot{\Psi}(0) = j^{-1}(ja) = a$$

Mientras, el momento central de segundo orden se calcula como,

$$\mu_{2\xi} = \sigma_{\xi}^2 = -\ddot{\Psi}(0) = -\frac{d^2}{d\omega^2} \left( j\omega a - \frac{b^2 \omega^2}{2} \right) \Big|_{\omega=0} = b^2$$

Así mismo, se calculan los momentos de tercer y cuarto orden, de acuerdo con la expresión (2.20):  $\mu_{3\xi} = 0$ ,  $\mu_{4\xi} = 3\mu_{2\xi}^2 = 3\sigma_{\xi}^4$ .



**Figura 2.4.** Distribución normal

La correspondiente función distributiva para un valor dado  $\xi_0$ , que se muestra en la Figura 2.4(b), se determina como:

$$F(\xi_0) = \int_{-\infty}^{\xi_0} \mathcal{N}(m_{1\xi}, \sigma_{\xi}) d\xi = \frac{1}{\sqrt{2\pi}\sigma_{\xi}} \int_{-\infty}^{\xi_0} \exp\left(-\frac{(\xi - m_{1\xi})^2}{2\sigma_{\xi}^2}\right) d\xi, \quad (2.24)$$

En la Figura 2.4 se muestran tanto la FDP como la distribución Gaussiana para diferentes valores de  $m_{1\xi_i}$  y  $\sigma_{\xi_i}$ ,  $i\{1, 2\}$ , en la que  $m_{1\xi_1} < m_{1\xi_2}$ , mientras,  $\sigma_{\xi_1} < \sigma_{\xi_2}$ .

En general, las siguientes son las razones por las cuales se emplea ampliamente la función densidad de probabilidad Gaussiana [14]:

1. Es un buen modelo matemático para la mayoría de los fenómenos aleatorios; es más, se puede demostrar que en muchos casos el modelo real converge al tipo Gaussiano (teorema del límite central).
2. La FDP normal es una de las pocas que puede describir fenómenos aleatorios compuestos por un número cualquiera de variables aleatorias.

3. Cualquier combinación lineal de variables aleatorias Gaussianas, también es Gaussiana.
4. Como se observa de (2.23), la FDP de Gauss se describe solo con dos parámetros: la media  $m_{1\xi}$  y la varianza  $\sigma_\xi^2$ .
5. El análisis matemático de los variables aleatorias muchas veces tiene solución solo para el caso Gaussiano.

Las principales propiedades de la FDP Gaussiana son las siguientes:

- (a). La curva normal es de forma simétrica con respecto a la media  $m_{1\xi}$ . En particular, a los valores de  $x_{a-} = m_{1\xi} - a$  y de  $x_{a+} = m_{1\xi} + a$  les corresponde un mismo valor de densidad.
- (b). El valor máximo de la curva está en  $\max(p_\xi(x)) = p_\xi(x = m_{1\xi}) = (\sigma_\xi \sqrt{2\pi})^{-1}$ , que es el valor de la variable aleatoria asignada e igual al primer momento inicial  $\xi$ . Dado  $\sigma_\xi = 1$ , entonces,  $\max(p_\xi(x)) = 0.39$ .
- (c). El valor de  $\sigma_\xi$  por la abscisa, que corresponde al punto de desdoblamiento de la curva, es  $0.6 \max(p_\xi(x))$ , con lo cual las 2/3 partes de todas las observaciones se distribuyen dentro del intervalo de los valores  $(m_{1\xi} - \sigma_\xi, m_{1\xi} + \sigma_\xi)$ .
- (d). Cualquier variable aleatoria con FDP Gaussiana, con alta probabilidad, toma valores cercanos a su esperanza, lo que se expresa por *la ley de los sigmas*:

$$P\{|\xi - m_{1\xi}| \geq k\sigma\} = \begin{cases} 0.3173, & k = 1 \\ 0.0455, & k = 2 \\ 0.0027, & k = 3 \end{cases} \quad (2.25)$$

Por cierto, tiene lugar la siguiente desigualdad:

$$P\{|\xi - m_{1\xi}| \geq k\sigma\} < 4/9k^2$$

que para  $k = 3$  implica que  $P\{|\xi - m_{1\xi}| \geq k\sigma\} < 4/81 \approx 0.05$ . En otras palabras, la probabilidad de desviación de la variable aleatoria del valor medio mayor a  $3\sigma_\xi$  es menor que 0.05.

- (e). En el caso de la ley de  $3\sigma$ , de la expresión (2.25) se pueden seleccionar los valores extremos que se consideran cumplen con la condición de estructura Gaussiana:

$$\xi_{\max} = m_{1\xi} + 3\sigma_\xi; \quad \xi_{\min} = m_{1\xi} - 3\sigma_\xi$$

La FDP *Gaussiana normalizada* se refiere al caso cuando la varianza  $\sigma_\xi^2$  es uno y la media es cero, esto es,  $(2\pi)^{-1/2} \exp(-\xi^2/2)$ . En este caso, la probabilidad de error para todos los

valores de la variable aleatoria  $\xi$  hasta  $x_0$  está dada por:

$$P\{\xi \leq x_0\} = \int_{-\infty}^{x_0} \mathcal{N}(0, 1) d\xi = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x_0} e^{-\frac{\xi^2}{2}} d\xi \triangleq \Phi(x_0), \quad (2.26)$$

Mediante el cambio de variables en (2.24), para la FDP Gaussiana se cumple que la probabilidad de ocurrencia de la variable aleatoria en el intervalo  $(x_1, x_2)$  es igual a:

$$P_\xi(x_1 \leq \xi < x_2) = \Phi\left(\frac{x_2 - m_{1\xi}}{\sigma_\xi}\right) - \Phi\left(\frac{x_1 - m_{1\xi}}{\sigma_\xi}\right)$$

En la práctica se tabula la función

$$\Phi_0(x_0) = \frac{1}{\sqrt{2\pi}} \int_0^{x_0} e^{-\frac{\zeta^2}{2}} d\zeta \quad (2.27)$$

conocida como la *Integral de Laplace* [1] y relacionada con  $\Phi(\xi)$  en (2.26) por la expresión,  $\Phi(\xi) = 1/2 + \Phi_0(\xi)$ . El cálculo de la integral, para valores pequeños de  $\xi$ , se realiza empleando la aproximación:

$$\Phi(\xi) = \frac{1}{2} + \frac{1}{\sqrt{2\pi}} \left( \xi - \frac{\xi^3}{2 \cdot 3} + \frac{\xi^5}{2! \cdot 2^2 \cdot 5} + \cdots + \frac{\xi^{2n+1}}{n! (2n+1) 2^n} + \cdots \right)$$

Otros ejemplos de modelos de FDP de amplio uso, para variables aleatorias continuas, son los siguientes:

### Densidad Gamma ( $\Gamma$ -densidad)

$$p_\xi(x; b, c) = \begin{cases} \frac{1}{b \Gamma(c)} \left(\frac{x-a}{b}\right)^{c-1} \exp\left(-\frac{x-a}{b}\right), & x \geq 0, b > 0, c \geq 0.5, x, a, b, c \in \mathbb{R} \\ 0, & \text{otros valores} \end{cases} \quad (2.28)$$

siendo  $x \in [0, \infty)$ , donde  $a$  se le conoce como el parámetro de posición, que generalmente se asume igual a cero,  $b > 0$  es de escala y  $c$  es el de forma.

La función Gamma  $\Gamma(n)$  ó función de Euler, que se define como,

$$\Gamma(n) = \int_0^\infty y^{n-1} \exp(-y) dy, \quad n > 0$$

Los momentos de orden  $k$  para la  $\Gamma$ -densidad están dados por la expresión

$$m_{k\xi} = \frac{c(c+1) \cdots (c+k-1)}{\lambda^k}, \quad \sigma_\xi^2 = c/\lambda^2$$



La  $\Gamma$ -densidad tiene importancia en la Teoría de Colas y describe el tiempo necesario para la aparición de  $c$  sucesos independientes que ocurren con igual intensidad  $\lambda$ .

---

**Ejemplo 2.7.** Un ferry comienza su ruta cada que se completan 9 automóviles, los cuales se consideran que llegan a la estación de manera independiente con una intensidad media de 6 unidades por hora. Determinar la probabilidad de que el tiempo entre los recorridos sucesivos del ferry sea menor que una hora.

Asumiendo que la variable aleatoria del tiempo entre las recorridos tiene  $\Gamma$ -densidad con parámetros:  $x = 1$ ,  $c = 9$ , y  $\lambda = 6$ , el cálculo de la respectiva probabilidad da como resultado  $P(1; 1/6, 9) = 0.153$ .

---

Como casos particulares de la  $\Gamma$ -densidad se consideran, entre otras, la FDP de *Erlang*, cuando  $c \in \mathbb{Z}^+$ , (entonces, se cumple que  $\Gamma(c) = (c-1)!$ ), la densidad  $\chi^2$  con  $n$  grados de libertad, cuando  $\lambda = 0.5$  y  $c = n/2$ . El caso especial, cuando  $c = 1.0$ , corresponde a la FDP *exponencial*, cuyo modelo se muestra en el ejemplo 2.8 y el cual se emplea en análisis de confiabilidad. Sin embargo, el parámetro  $\lambda$  asociado al tiempo medio entre fallas se asume constante, generando una fuerte restricción en el modelo exponencial.

Si se considera que  $\lambda(t) = \text{var}$  en el tiempo, particularmente,  $\lambda(t) = c/b((x-a)/b)^{c-1}$ , el modelo resultante de FDP corresponde a la densidad de *Weibull*.

$$p_{\xi}(x; a, b, c) = c/b((x-a)/b)^{c-1} \exp(-((x-a)/b)^c), \quad x \geq a, b > 0, c > 0 \quad (2.29)$$

La intensidad de fallas y la respectiva FDP toman diferentes formas variando el parámetro  $c$ . Particularmente, cuando  $c < 1$ , la FDP (2.29) es una curva monótona decreciente y con el transcurso del tiempo la intensidad de las fallas disminuye. Para  $c = 1$ , la intensidad de fallas es constante y la FDP de Weibull coincide con la exponencial, mientras para  $c > 1$  la densidad tiene forma de pico y la intensidad de fallas aumenta con el tiempo.

**Densidad de Student ( $t$ -densidad).** Corresponde a la FDP de la variable:

$$t = \frac{\xi_0}{\sqrt{1/n \left( \sum_{k=1}^n \xi_k^2 \right)}}$$

donde  $\xi_k$  son valores independientes, tales que  $\{\xi_k \in \mathcal{N}(0, 1) : k = 0, \dots, n\}$ . La densidad de la variable aleatoria  $t$  es igual a,

$$p_{\xi}(x; n) = \frac{\Gamma((n+1)/2)}{\sqrt{\pi n} \Gamma(n/2)} \left(1 + x^2/n\right)^{-(n+1)/2}, \quad -\infty < x < \infty \quad (2.30)$$

En variadas aplicaciones, se tiene que  $n \in \mathbb{N}$ , así, para  $n = 1$ , la función (2.30) corresponde a la FDP de *Cauchy* (que no tiene momentos ni de primer ni de segundo orden!):

$$p_{\xi}(x, \lambda) = \frac{1}{\pi} \frac{\lambda}{\lambda^2 + (x - n)}, \quad \lambda > 0$$

La forma de la  $t$ -densidad es similar a la curva de Gauss (para  $n \geq 40$  converge en la normal), no obstante, en lugar de depender de la media y la varianza, se expresa solamente en función del valor  $n$  ó *grados de libertad*. La principal aplicación de la  $t$ -densidad es la prueba de hipótesis para los valores medios.

**$\chi^2$ - densidad.** Sea la variable aleatoria  $\chi^2 = \sum_{k=1}^n \xi_k^2$ , conformada por los valores independientes  $\xi_k^2$ ,  $\{\xi_k \in \mathcal{N}(0, 1)\}$ , entonces la FDP con  $n$  grados de libertad:

$$p_{\xi}(x; n) = \begin{cases} \frac{1}{2^{\frac{n}{2}} \Gamma(\frac{n}{2})} x^{\frac{n}{2}-1} \exp(-\frac{x}{2}), & x > 0 \\ 0, & x \leq 0 \end{cases} \quad (2.31)$$

En la Figura 2.5, se muestran ejemplos para diferentes valores de  $n$  de la densidad  $\chi^2$ , la cual corresponde al caso particular de la  $\Gamma$ -densidad cuando  $b = 2$  y  $c = k/2$ . Esta FDP se determina sobre el dominio de valores  $0 \leq \xi < \infty$  y le corresponden los momentos de orden  $k$ :

$$m_{k\xi} = n(n+2) \cdots (n+2(k-1))$$

con varianza  $\sigma_{\chi^2}^2 = 2n$ .

Al aumentar el valor de  $n$ , la  $\chi^2$ -densidad tiende a la FDP normal, pero con menor velocidad que la  $t$ -densidad, por lo que para valores grandes de  $x$  y  $n$ , se demuestra que

$$F_{\chi^2(n)} \approx \Phi\left(\sqrt{2x} - \sqrt{2n-1}\right)$$

La  $\chi^2$ -densidad cumple el principio de aditividad,  $\chi^2(n_1 + n_2) = \chi^2(n_1) + \chi^2(n_2)$ .

El área principal de aplicación de esta FDP es la prueba de hipótesis. Cuando el número  $n$  de variables aleatorias a considerar es igual a 3:  $\chi^2 = \xi_1^2 + \xi_2^2 + \xi_3^2$ , asumiendo su estructura Gaussiana y su total independencia mutua, la respectiva FDP se denomina de *Maxwell*, mientras, para  $n = 2$ , corresponde a la FDP de *Rayleigh* y cuando es  $n = 1$ , se obtiene la FDP de la potencia de la variable original  $\xi$ . La densidad de Rayleigh también corresponde al caso particular de la FDP de Weibull, cuando  $a = 0$  y  $c = 2$ .

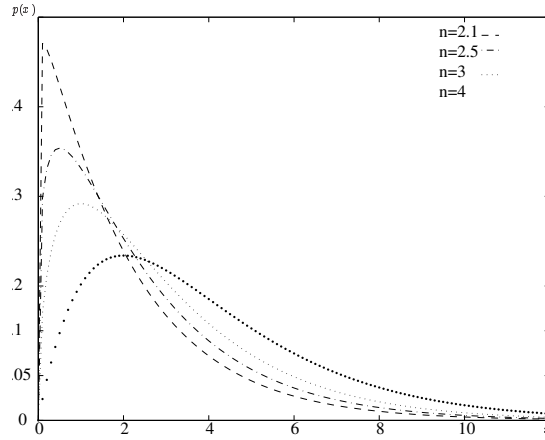


Figura 2.5. Densidad  $\chi^2$

**FDP de Fisher ( $F$ - densidad).** Corresponde a la FDP de la variable aleatoria

$$\xi = \left(\eta_{n_1/n_1}\right) / \left(\eta_{n_2/n_2}\right) \quad (2.32)$$

donde  $\eta_{n_i}$  son valores aleatorios independientes con  $\chi^2$ -distribución y  $n_i$ ,  $i = 1, 2$  grados de libertad. Un ejemplo particular, corresponde al caso cuando se tienen dos conjuntos de observaciones: el primero,  $\{x_i : i = 1, \dots, m_1\}$  y segundo,  $\{y_i : i = 1, \dots, m_2\}$ , ambas con FDP Gaussiana  $\mathcal{N}(m_{1\eta}, \sigma_\eta^2)$ ,  $\eta \in \{x, y\}$ . Entonces, la estadística

$$\frac{1}{m_1 - 1} \sum_{i=1}^{m_1} (x_i - \tilde{m}_{1x}) \bigg/ \frac{1}{m_2 - 1} \sum_{i=1}^{m_2} (y_i - \tilde{m}_{1y})$$

siendo  $\tilde{m}_{1\eta} = m_{1\eta}^{-1} \sum_{i=1}^{m_{1\eta}} \eta_i$ ,  $\eta \in \{x, y\}$ , le corresponde la  $F$ -densidad con  $(m_1, m_2)$  grados de libertad.

De otra parte, si se tiene la variable aleatoria  $\eta$  descrita por la  $F$ -densidad con  $(m_1, m_2)$  grados de libertad, entonces la variable aleatoria  $\xi = \frac{1}{2} \ln \eta$  le corresponde la  $z$ -densidad con  $(m_1, m_2)$  grados de libertad, para la cual se tienen los siguientes momentos:

$$m_{1\xi} = 0, \quad \sigma_\xi^2 = \frac{1}{2} \frac{m_1 + m_2}{m_1 m_2}$$

En general, la  $F$ -densidad se determina para el dominio  $0 \leq \xi < \infty$  y tiene aplicación en el análisis de varianza, la prueba múltiple de hipótesis y el análisis de regresión.

**Densidad logonormal.** Sea el par de variables aleatorias relacionadas por la transformación:  $\xi = \ln \eta$ , donde la variable aleatoria  $\xi$  es de estructura Gaussiana  $\mathcal{N}(m_{1\xi}, \sigma_\xi^2)$ , entonces se puede demostrar que la FDP para  $\eta$  tiene la forma

$$p_\eta(x) = \begin{cases} \frac{1}{x \sigma_\xi \sqrt{2\pi}} \exp\left(-\frac{(\ln x - m_{1\xi})^2}{2\sigma_\xi^2}\right), & x > 0 \\ 0, & x \leq 0 \end{cases}$$

con momentos,

$$m_{k\eta} = \exp\left(\frac{1}{2} k^2 \sigma_\xi^2 + k m_{1\xi}\right) \text{ y } \sigma_\eta^2 = \left(\exp(\sigma_\xi^2) - 1\right) \exp(\sigma_\xi^2 + 2m_{1\xi})$$

Si  $\eta$  es del tipo  $\mathcal{N}(0, 1)$ , entonces la variable  $\xi = \exp(\sigma_\xi \eta + m_{1\xi})$  tiene FDP logonormal con parámetros  $(m_{1\xi}, \sigma_\xi)$ . La FDP logonormal encuentra aplicación en tareas estadísticas relacionadas con mediciones experimentales de la física, geología, economía, biología, etc.

**Densidad uniforme.** Corresponde al caso en que todos los valores de la variable aleatoria tienen igual probabilidad de ocurrencia:

$$p_\xi(x) = \begin{cases} \frac{1}{b-a}, & x \in [a, b] \\ 0, & x \notin [a, b] \end{cases} \quad (2.33)$$

cuya función de distribución tiene la forma:

$$F_{\xi}(x) = \begin{cases} 0, & x < a \\ (x - a)/(b - a), & x \in [a, b] \\ 1, & x > b \end{cases}$$

La densidad uniforme le corresponden los momentos:

$$m_{k\xi} = \frac{b^{k+1} - a^{k+1}}{(b - a)(k + 1)}$$

con varianza  $\sigma_{\xi}^2 = (b - a)^2 / 12$ .

Al realizar la transformación lineal  $\eta = (\xi - a)/(b - a)$  se llega a la FDP uniforme en el intervalo  $[0, 1]$ . Esta propiedad explica su amplia aplicación en los métodos de simulación, por ejemplo el de Monte Carlo.

**Ejemplo 2.8.** Hallar los valores de la media y la varianza de las siguientes FDP:

**Exponencial.** Dada por el modelo de FDP (Figura 2.6a)

$$p_{\xi}(xi) = \begin{cases} \lambda e^{-\lambda x}, & x \geq 0 \\ 0, & x < 0 \end{cases} \quad (1)$$

a la cual le corresponden los valores de media y varianza, respectivamente,

$$m_{1\xi} = \lambda \int_0^{\infty} x e^{-\lambda x} dx = \frac{1}{\lambda}, \quad \sigma_{\xi}^2 = \lambda \int_0^{\infty} x^2 e^{-\lambda x} dx - \frac{1}{\lambda^2} = \frac{2}{\lambda^2} - \frac{1}{\lambda^2} = \frac{1}{\lambda^2}$$

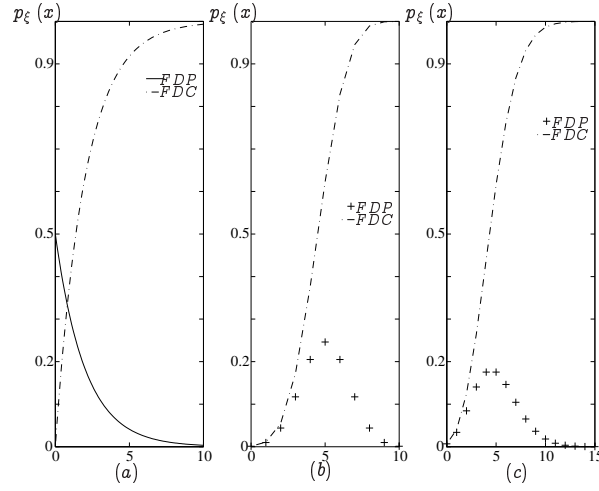
**Binomial.** Descrita por el modelo (Figura 2.6b):

$$p_k = P\{\xi = k\} = \begin{cases} 0, & k < 0 \\ C_n^k a^k (1 - a)^{n-k}, & k \leq n \\ 0, & k > n \end{cases} \quad (2)$$

con primer momento inicial:  $m_{1\xi} = \sum_{k=0}^n k C_n^k a^k (1 - a)^{n-k} = na$ , y segundo momento centralizado  $\sigma_{\xi}^2 = \sum_{k=0}^n k^2 C_n^k a^k (1 - a)^{n-k} = na(1 - a)$ .

**Poisson.** Para el caso asintótico de la FDP binomial, obtenido cuando  $n \rightarrow \infty$ , y  $a \rightarrow 0$ , siendo  $na = \lambda = \text{const.}$ , (el parámetro  $\lambda$  se denomina *valor promedio de ocurrencia* del evento), se puede demostrar que esta densidad se convierte en (Figura 2.6c):

$$p_k = \frac{\lambda^k}{k!} e^{-\lambda} u(k), \quad (3)$$



**Figura 2.6.** Ejemplos de distribuciones de probabilidad

A la densidad (3) le corresponden los siguientes momentos:

$$m_{1\xi} = \sum_{k=0}^{\infty} k \frac{\lambda^k}{k!} e^{-\lambda} = \lambda,$$

$$\sigma_{\xi}^2 = \sum_{k=0}^{\infty} k^2 \frac{\lambda^k}{k!} e^{-\lambda} = \sum_{k=0}^{\infty} k \frac{\lambda^k}{k!} e^{-\lambda} + \sum_{k=0}^{\infty} k(k-1) \frac{\lambda^k}{k!} e^{-\lambda} = \lambda$$

En la Figura 2.7 se muestran las relaciones que existen entre las diferentes densidades analizadas de probabilidad, en la cual algunos enlaces corresponden a simplificación o generalización de los modelos de aleatoriedad, o bien al empleo de la transformación  $\mathcal{G}\{\cdot\}$ . En líneas punteadas están indicadas las FDP discretas.

**Transformación de modelos de FDP.** Sea la variable aleatoria  $x \in \xi$  que se procesa en un dispositivo, cuya salida está dada por la relación  $y = f(x)$ ,  $y \in \eta$ . La transformación de la variable aleatoria se obtiene como:

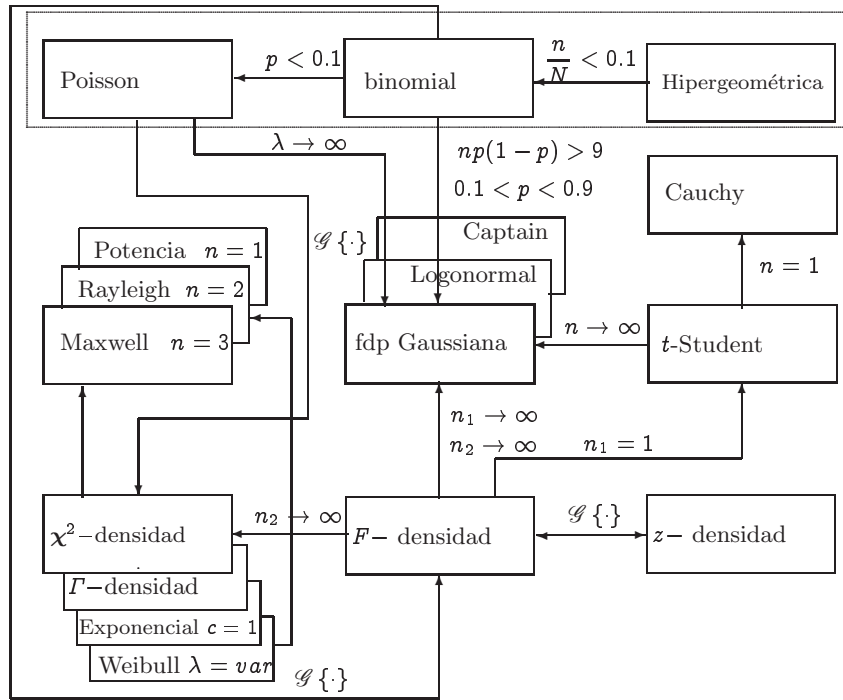
$$P\{\xi \leq x\} = P\{\xi \leq f^{-1}(y)\} = P\{\eta \leq y\}$$

además,

$$F_{\xi}(y) = P\{\eta \leq y\} = P\{f(x) \leq y\} = P\{x \leq f^{-1}(y)\} = F_{\xi}\{f^{-1}(y)\}$$

entonces, finalmente se obtiene

$$\frac{dF_{\eta}(y)}{dy} = \frac{dF_{\eta}(y)}{dx} \frac{dx}{dy} = \frac{dF_{\xi}(f^{-1}(y))}{dx} \frac{dx}{dy}.$$



**Figura 2.7.** Relación entre diferentes FDP.  $\mathcal{G}\{\cdot\}$  transformación funcional

Como es conocido, se debe cumplir que  $p(\xi) \geq 0$ , entonces, la definición de la FDP a la salida del dispositivo, expresada con respecto a las características de la variable aleatoria a la entrada, se ajusta de la siguiente manera:

$$p_{\eta}(y) = p_{\xi}(x) \frac{|dx|}{|dy|}, \quad (2.34)$$

donde el empleo de valores absolutos en (2.34) se hace para asegurar que en ningún momento la FDP sea negativa.

Cabe anotar, que el cálculo de los momentos después de la transformación se puede realizar sin necesidad de conocer la FDP de salida, en particular,

$$E\{(\eta - m_{1\eta})^n\} = \int_{-\infty}^{\infty} (y - m_{1\eta})^n p_{\eta}(y) dy = \int_{-\infty}^{\infty} (f(x) - m_{1\eta})^n p_{\xi}(x) dx \quad (2.35)$$

**Ejemplo 2.9.** Sea  $\eta = f(\xi) = 1 - e^{-\xi}$ ,  $0 \leq \xi \leq 2$ ,  $p(\xi) = ce^{-\xi}$ . Hallar el respectivo valor de  $c$  y la correspondiente FDP de salida,  $p(\eta)$ .

Por cuanto,  $\int_0^2 p(\xi) d\xi = c \int_0^2 e^{-\xi} d\xi = -ce^{-\xi} \Big|_0^2 = (-ce^{-2} + c) = 1$ .

Entonces,  $c = 1/0.865 = 1.156$ . Además,  $\xi = f^{-1}(\eta) = -\ln(1 - \eta)$ ,  $\frac{|d\eta|}{|d\xi|} = 1/(1 - \eta)$ ,

luego,  $p(\eta) = ce^{\ln(1-\eta)} 1 / (1 - \eta) = c = \frac{1}{\eta_{\max} - \eta_{\min}} = 1.156$ .

De lo anterior se deduce que la FDP  $p(\eta)$  corresponde al modelo uniforme.

**Ejemplo 2.10.** Hallar el valor medio y la varianza de la variable aleatoria de salida  $y \in \eta$ , dados los momentos  $m_{1\xi}$  y  $\mu_{2\xi}$  de la variable aleatoria entrada  $x \in \xi$ ; ambas variables relacionadas por la dependencia lineal  $y = ax + b$ .

De acuerdo con la expresión (2.35), se tiene que

$$\begin{aligned} E\{y\} &= \int_{-\infty}^{\infty} (ax + b) p_{\xi}(x) dx = a \int_{-\infty}^{\infty} x p_{\xi}(x) dx + b \int_{-\infty}^{\infty} p_{\xi}(x) dx \\ &= am_{1\xi} + b \\ E\{(y - m_{1\eta})^2\} &= \int_{-\infty}^{\infty} (ax + b - m_{1\eta})^2 p_{\eta}(x) dx = \int_{-\infty}^{\infty} (ax - b - am_{1\xi} - b)^2 p_{\xi}(x) dx \\ &= a^2 \int_{-\infty}^{\infty} (x - m_{1\xi})^2 p_{\xi}(x) dx = a^2 \sigma_{\xi}^2 \end{aligned}$$

**Descomposición de FDP.** En el análisis y transformación de variables aleatorias, puede ser de ayuda la descomposición ortogonal (1.4) de las diferentes FDP:

$$p_{\xi}(x) = \sum_{k=-\infty}^{\infty} x_k w_p(x) \phi_k(x) \quad (2.36)$$

siendo  $x_k$  los coeficientes de descomposición (1.7),  $\phi_k(x)$  la base ortogonal de descomposición y  $w_p(x)$  una función ventana de ajuste, que corresponde a la versión normalizada de la función de peso  $w(x)$ , definida en (1.9). En la práctica, en calidad de función ventana se toma una función de probabilidad conocida, por ejemplo la FDP normalizada,  $\mathcal{N}(0, 1)$ . Por cuanto el objetivo de la representación es la aproximación de las funciones de probabilidad, de tal manera que se pueda brindar la precisión deseada con el menor número de elementos de la serie en (2.36), entonces se emplean polinomios ortogonales, por ejemplo, los de *Chebyshev*, *Hermite*, *Laguerre*, *Legendre*, entre otros [3]

**Ejemplo 2.11.** Sea una FDP simétrica,  $p_{\xi}(x) = p_{\xi}(-x)$ , definida sobre el intervalo cerrado  $[-\pi \leq x \leq \pi]$ , e igual a  $p_{\xi}(x) = 0$  para todos los valores de  $x$  fuera del intervalo. Hallar la descomposición de la respectiva varianza por la serie de Fourier.

La descomposición por la serie (1.8), teniendo en cuenta la paridad de la FDP, se reduce solo a los términos reales,

$$p_{\xi}(x) = \frac{a_0}{2} + \sum_{n=1}^{\infty} a_n \cos nx$$

donde  $a_0$  se halla de la condición de unidad de área para  $p_\xi(x)$ ,

$$\int_{-\infty}^{\infty} \left( \frac{a_0}{2} + \sum_{n=1}^{\infty} a_n \cos nx \right) dx = 1$$

$$\text{entonces, } \frac{a_0}{2} x \Big|_{-\pi}^{\pi} + \sum_{n=1}^{\infty} \frac{a_n}{n} \sin nx \Big|_{-\pi}^{\pi} = 1.$$

Luego,  $a_0 = 1/\pi$ , y por lo tanto, la descomposición de la FDP, mediante la serie de Fourier, toma la forma

$$p_\xi(x) = \frac{1}{2\pi} + \sum_{n=1}^{\infty} a_n \cos nx$$

De la propiedad (2.13), se sabe que  $\sigma_\xi^2 = 2m_{1\xi}$ , con lo cual se obtiene el siguiente valor de varianza:

$$\sigma_\xi^2 = 2 \int_{-\infty}^{\infty} x^2 \left( \frac{1}{2\pi} + \sum_{n=1}^{\infty} a_n \cos nx \right) dx = \frac{\pi^2}{3} + 4\pi \sum_{n=1}^{\infty} (-1)^n \frac{a_n}{n^2}$$

En la práctica, se realizan las siguientes presunciones sobre la FDP a representar:

- (a). las funciones son unimodales, en el sentido en que presentan un solo máximo,
- (b). Las FDP presentan ramales por ambos lados del máximo, que caen suficientemente rápido hasta cero, en la medida en que aumenta el valor absoluto del argumento.

Un ejemplo corresponde a la descomposición mediante los polinomios de Hermite:

$$p_\xi(x) = w_p(x) \sum_{k=-\infty}^{\infty} \frac{x_k}{\sqrt{k!}} H_k(x) \quad (2.37)$$

para la cual, se tiene la ventana  $w(x) = e^{-x^2}$ , entonces  $w_p(x) = \mathcal{N}(0, 1)$ . Los coeficientes

$$x_k = \frac{1}{\sqrt{k!}} \int_{-\infty}^{\infty} p_\xi(x) H_k(x) dx = \frac{1}{\sqrt{k!}} E\{H_k(x)\}$$

son denominados *cuasimomentos* [15]. Donde los valores  $x_0 = 0$  y  $x_1 = x_2 = 0$ , debido a la condición de normalización de la función de densidad de probabilidad.

Si en la descomposición (2.37), la serie polinomial se limita hasta los primeros  $N$  términos, entonces se obtiene la *serie de Edworth*:

$$p_\xi(x) \approx w_p(x) \left( 1 + \sum_{n=3}^N \frac{x_n}{\sqrt{n!}} H_n(x) \right) = w_p(x) \left( 1 + \frac{\gamma_{1\xi}}{3!} H_3(x) + \frac{\gamma_{2\xi}}{4!} H_4(x) \right)$$

siendo  $\gamma_{i\xi}$ ,  $i = 1, 2$ , los momentos definidos en (2.15) y (2.16), respectivamente.



#### 2.1.4. Modelos de funciones de probabilidad con dimensión múltiple

Sea un fenómeno  $\xi$  descrito por  $n \geq 2$  variables aleatorias reales  $\{x_k \in \xi : k = 1, \dots, n\}$ , o *proceso aleatorio con dimensión múltiple*, para el cual los valores medios y momentos se encuentran empleando la integración múltiple por cada variable, incluyendo la función de densidad conjunta que también es de dimensión  $n$ .

**Modelos de probabilidad de dos dimensiones.** En el caso particular de dos dimensiones, la función de distribución acumulativa conjunta corresponde a la probabilidad de ocurrencia simultánea de las desigualdades  $x_1 \leq \lambda_1$ ,  $x_2 \leq \lambda_2$ :

$$F_\xi(\lambda_1, \lambda_2) = P\{x_1 \leq \lambda_1, x_2 \leq \lambda_2\} = \int_{-\infty}^{\lambda_1} \int_{-\infty}^{\lambda_2} p_\xi(x_1, x_2) dx_1 dx_2, \quad \lambda_1, \lambda_2 \in \mathbb{R} \quad (2.38)$$

Las siguientes propiedades se cumplen para las funciones de probabilidad de dos dimensiones:

- (a).  $0 \leq F_\xi(\lambda_1, \lambda_2) \leq 1$ . Por cierto,  $F_\xi(\lambda_1, -\infty) = F_\xi(-\infty, \lambda_2) = F_\xi(-\infty, -\infty) = 0$ , mientras, el otro caso extremo tiene valor  $F_\xi(\infty, \infty) = 1$
- (b).  $\frac{\partial}{\partial x_i} F_\xi(\lambda_1, \lambda_2) \geq 0$ ,  $i = 1, 2$ . De lo cual resulta que  $F_\xi(\lambda_1, x_2) \geq F_\xi(\lambda_2, x_2)$ , asumiendo que  $\lambda_2 > \lambda_1$ ;  $\lambda_1, \lambda_2 \in x_1$ , o bien,  $F_\xi(x_1, \lambda_1) \geq F_\xi(x_1, \lambda_2)$ ,  $\forall \lambda_2 > \lambda_1$ ;  $\lambda_1, \lambda_2 \in x_2$ .
- (c).  $p_\xi(x_1, x_2) = \frac{\partial^2}{\partial x_1 \partial x_2} F_\xi(x_1, x_2)$ ,
- (d).  $p_\xi(x_1, x_2) \geq 0$ ,

Las funciones de probabilidad de dimensión simple se expresan a través de las respectivas funciones de dos dimensiones, de la siguiente manera:

$$\begin{cases} F_\xi(\lambda_1) = \lim_{\lambda_2 \rightarrow \infty} F_\xi(\lambda_1, \lambda_2) = \int_{-\infty}^{\lambda_1} \int_{-\infty}^{\infty} p_\xi(x_1, x_2) dx_1 dx_2 \\ p_\xi(x_1) = \frac{\partial F_\xi(x_1)}{\partial x_1} = \int_{-\infty}^{\infty} p_\xi(x_1, x_2) dx_2, \end{cases} \quad (2.39)$$

De forma similar a lo obtenido en (2.22), para los valores suficientemente pequeños de  $\Delta_1 \in x_1$  y  $\Delta_2 \in x_2$ , tiene lugar la aproximación:

$$F_\xi(\lambda_1 \leq x_1 < \lambda_1 + \Delta_1, \lambda_2 \leq x_2 < \lambda_2 + \Delta_2) \approx p_\xi(x_1, x_2) \Delta_1 \Delta_2$$

Cuando las variables aleatorias  $x_1$  y  $x_2$ , que pertenecen al fenómeno  $\xi$ , son estadísticamente independientes, la condición necesaria y suficiente se expresa por medio de la relación

común con la FDP conjunta:

$$p_{\xi}(x_1, x_2) = p_{\xi}(x_1) p_{\xi}(x_2) \quad (2.40)$$

De otra parte, el nivel de dependencia estadística entre dos diferentes variables de un fenómeno aleatorio continuo aleatorio  $\xi$  está dado por la FDP condicional

$$\begin{cases} p_{\xi}(x_1 | x_2) = \frac{p_{\xi}(x_1, x_2)}{p_{\xi}(x_2)} = \frac{p_{\xi}(x_1, x_2)}{\int_{-\infty}^{\infty} p_{\xi}(x_1, x_2) dx_2} \\ p_{\xi}(x_1 | x_2) = \frac{p_{\xi}(x_1, x_2)}{p_{\xi}(x_1)} = \frac{p_{\xi}(x_1, x_2)}{\int_{-\infty}^{\infty} p_{\xi}(x_1, x_2) dx_1} \end{cases} \quad (2.41)$$

**Momentos de correlación.** Además de los valores de aleatoriedad definidos para variables con dimensión simple, en el caso particular de variables múltiples con dos dimensiones, se define el momento conjunto de primer orden [16]:

$$E\{g(x_1, x_2)\} \triangleq \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x_1, x_2) p_{\xi_1 \xi_2}(x_1, x_2) dx_1 dx_2 \quad (2.42)$$

siendo  $g(x_1, x_2)$  la *función generatriz* del momento conjunto.

La función generatriz más común corresponde al caso en que los momentos y los cumulantes de las variables aleatorias con dimensión múltiple se determinan como los coeficientes de descomposición en la serie exponencial de las respectivas funciones características o de sus logaritmos, por lo tanto similar al caso de análisis de una dimensión, los primeros coeficientes de descomposición son los más importantes y significativos.

Sea la función  $g(x_1, x_2)$ , que contiene las variables aleatorias  $\xi_1$  y  $\xi_2$ , cuyo valor medio se determina por (2.42). La descomposición de la función  $g(x_1, x_2)$  mediante la serie de Taylor en la vecindad de los puntos  $(m_{1\xi_1}, m_{1\xi_2})$ , donde se supone ocurre la mejor estimación de la función original, tiene la forma:

$$g(x_1, x_2) = g(m_{1\xi_2}, m_{1\xi_2}) + \frac{\partial g}{\partial x_1}(x_1 - m_{1\xi_1}) + \frac{\partial g}{\partial x_2}(x_2 - m_{1\xi_2}) + \dots$$

Si la FDP con dos dimensiones  $p_{\xi_1 \xi_2}(x_1, x_2)$  está concentrada en la cercanía de los puntos  $(\xi_1, \xi_2)$  y cambia suavemente en esta vecindad en proporción a  $(\sigma_{\xi_1}, \sigma_{\xi_2})$ , entonces la descomposición obtenida por Taylor, se puede limitar hasta los dos primeros términos:

$$E\{g(x_1, x_2)\} \simeq g(m_{1\xi_2}, m_{1\xi_2}) + \frac{1}{2} \left( \sigma_{\xi_1}^2 \frac{\partial^2 g}{\partial x_1^2} + 2\mu_{11} \frac{\partial^2 g}{\partial x_1 \partial x_2} + \sigma_{\xi_2}^2 \frac{\partial^2 g}{\partial x_2^2} \right) \quad (2.43)$$

siendo

$$\mu_{11} = E\{(\xi_1 - m_{1\xi_1})(\xi_2 - m_{1\xi_2})\} \triangleq K_{\xi_1\xi_2} \quad (2.44)$$

El valor del momento conjunto central de segundo orden  $\mu_{11}$  se denomina *momento* o *función de covarianza* conjunta entre las variables  $\xi_m$  y  $\xi_n$ , la cual se puede asimilar con el grado de dispersión o varianza mutua entre el par de variables aleatorias de dimensión simple. La expresión (2.43) muestra el papel importante que cumplen los momentos de menor orden, en particular, la esperanza, la varianza y el momento de correlación.

Se observa de (2.44) que la función generatriz es de la forma  $g(x_1, x_2) = x_1 x_2$ , por lo tanto, se define el momento inicial conjunto entre las variables  $\xi_m$  y  $\xi_n$  o *función de correlación*:

$$m_{11} = E\{\xi_1 \xi_2\} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x_1 x_2 p_{\xi_1, \xi_2}(x_1, x_2) dx_1 dx_2 \triangleq R_{\xi_1 \xi_2} \quad (2.45)$$

Al calcular los promedios conjuntos de dos variables, sin tener en cuenta sus respectivas escalas ni valores medios (obtenidos para el caso de dimensión simple), se emplea el *índice de correlación* o función de covarianza normalizada, que se define como:

$$\begin{aligned} \rho &\triangleq E\{(\xi_1 - m_{1\xi_1})/\sigma_{\xi_1} (\xi_2 - m_{1\xi_2})/\sigma_{\xi_2}\} \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{(x_1 - m_{1\xi_1})}{\sigma_{\xi_1}} \frac{(x_2 - m_{1\xi_2})}{\sigma_{\xi_2}} p_{\xi_1, \xi_2}(x_1, x_2) dx_1 dx_2 \end{aligned} \quad (2.46)$$

Entre las principales propiedades del coeficiente de correlación están las siguientes:

- (a).  $\rho = E\{\xi'_m \xi'_n\}$ , siendo  $\xi'_k = (\xi_k - m_{1\xi_k})/\sigma_{\xi_k}$ ,  $k = \{m, n\}$  la forma normalizada estadísticamente de la variable  $\xi_k$ , de tal manera que  $E\{\xi'_k\} = 0$  y  $\sigma_{\xi'_k}^2 = 1$ .
- (b).  $-1 \leq \rho \leq 1$ , por cuanto,

$$\begin{aligned} E\{(\xi'_m \pm \xi'_n)^2\} &= E\{\xi'^2_m \pm 2\xi'_m \xi'_n + \xi'^2_n\} = 1 \pm 2\rho + 1 \\ &= 2(1 \pm \rho) \geq 0 \end{aligned}$$

A partir de la definición 2.5, se tiene que un valor 0 para la función de correlación (2.45), implica que ambas variables,  $\xi_m$  y  $\xi_n$ , son ortogonales entre sí. Si la respectiva función de covarianza (2.44) es cero, entonces se dice que las variables  $\xi_m$  y  $\xi_n$  son linealmente independientes. Por cierto, de la definición (2.44) se deduce que dos variables aleatorias ortogonales cumplen la condición de independencia lineal si sus respectivos valores medios son iguales a cero. Por último, se tiene un tipo de independencia más fuerte cuando las variables aleatorias  $\xi_m$  y  $\xi_n$  son estadísticamente independientes, entonces se cumple que  $\rho = E\{\xi'_m \xi'_n\} = 0$ . En este caso, la FDP conjunta corresponde a la expresión (2.40), de la cual resulta que,  $E\{\xi_m, \xi_n\} = m_{1\xi_m} m_{1\xi_n}$ , por lo que la independencia estadística implica

la independencia lineal. El caso contrario no siempre es cierto, por cuanto la independencia estadística es una condición más fuerte que la independencia lineal.

---

**Ejemplo 2.12.** Hallar los momentos  $\mu_{11}$  y  $\sigma_{\xi \pm \eta}^2$  de las variables dependientes  $\eta$  y  $\xi$ .

Teniendo en cuenta la expresión (2.44), se tiene

$$\begin{aligned}\mu_{11} &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x_1 - m_{1\xi_1})(x_2 - m_{1\xi_2}) p_{\xi_1, \xi_2}(x_1, x_2) dx_1 dx_2 \\ &= E\{\xi_1 \xi_2\} - m_{1\xi_1} E\{\xi_2\} - m_{1\xi_2} E\{\xi_1\} + m_{1\xi_1} m_{1\xi_2} \\ &= E\{\xi_1, \xi_2\} - m_{1\xi_1} m_{1\xi_2}\end{aligned}$$

De la definición de la varianza se tiene,

$$\begin{aligned}\sigma_{\xi \pm \eta}^2 &= E\{(\xi \pm \eta)^2\} - (E\{\xi \pm \eta\})^2 = E\{\xi^2 \pm 2\xi\eta + \eta^2\} - (E\{\xi\} \pm E\{\eta\})^2 \\ &= E\{\xi^2\} \pm E\{2\xi\eta\} + E\{\eta^2\} - E^2\{\xi\} \mp 2E\{\xi\}E\{\eta\} - E^2\{\eta\} \\ &= E\{\xi^2\} - E^2\{\xi\} + E\{\eta^2\} - E^2\{\eta\} \pm E\{2\xi\eta\} \\ &= \sigma_{\xi_1}^2 + \sigma_{\xi_2}^2 \pm E\{2\xi\eta\}\end{aligned}$$

El anterior resultado se puede generalizar para la suma de  $m$  variables dependientes:

$$\sigma^2 \left( \sum_{l=1}^m \xi_l \right) = \sum_{l=1}^m \sum_{n=1}^m K_{nl}$$


---

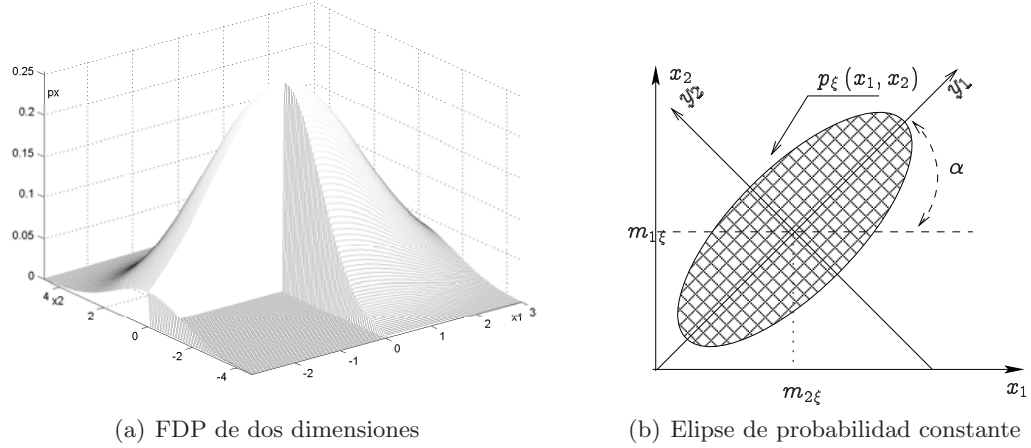
**Variables conjuntas Gaussianas.** Por definición, dos variables aleatorias  $\xi_1$  y  $\xi_2$  se denominan *conjuntamente Gaussianas* si su FDP conjunta tiene la forma:

$$p_{\xi_1 \xi_2}(x_1, x_2) = k \exp \left( - \left( ax_1^2 + bx_1x_2 + cx_2^2 + dx_1 + ex_2 \right) \right), \quad k \in \mathbb{R}$$

tal que la forma cuadrática de la exponencial,  $(ax_1^2 + bx_1x_2 + cx_2^2 + dx_1 + ex_2) \geq 0$ ,  $\forall x_1$  y  $x_2$ . La FDP Gaussiana conjunta expresada en función de sus momentos toma la forma:

$$\begin{aligned}p_{\xi_1 \xi_2}(x_1, x_2) &= \frac{1}{2\pi\sigma_{\xi_1}\sigma_{\xi_2}\sqrt{1-\rho^2}} \times \\ &\times \exp \left( \frac{-\sigma_{\xi_2}^2(x_1 - m_{1\xi})^2 + 2\rho\sigma_{\xi_1}\sigma_{\xi_2}(x_1 - m_{1\xi})(x_2 - m_{2\xi}) - \sigma_{\xi_1}^2(x_2 - m_{2\xi})^2}{2\sigma_{\xi_1}^2\sigma_{\xi_2}^2(1-\rho^2)} \right)\end{aligned}\tag{2.47}$$

Las siguientes son las propiedades de la FDP Gaussiana de 2 dimensiones (Figura 2.8(a)), las cuales se pueden generalizar a casos con dimensión múltiple de mayor orden.



**Figura 2.8.** Distribución de Gauss de dos dimensiones

- (a). Si dos variables aleatorias  $\xi_1$  y  $\xi_2$  no están correlacionadas, esto es,  $\rho = 0$ , entonces de (2.47) resulta que su FDP conjunta es igual a la multiplicación de la densidad de probabilidad de cada una de las variables:  $p_{\xi_1 \xi_2}(x_1, x_2) = p_{\xi_1}(x_1)p_{\xi_2}(x_2)$ . De acuerdo con (2.40), las variables son estadísticamente independientes entre sí.
- (b). Dos variables aleatorias,  $\xi_1$  y  $\xi_2$ , conjuntas Gaussianas dependientes (con valor de correlación  $\rho \neq 0$ ), siempre se pueden expresar en forma de dos nuevas variables aleatorias,  $\eta_1$  y  $\eta_2$ , que no tengan correlación (independientes), usando la siguiente transformación lineal (elipse de probabilidad constante de la Figura 2.8(b)):

$$\eta_1 = (\xi_1 - m_{1\xi_1}) \cos \alpha + (\xi_2 - m_{1\xi_2}) \sin \alpha, \quad \eta_2 = -(\xi_1 - m_{1\xi_1}) \sin \alpha + (\xi_2 - m_{1\xi_2}) \cos \alpha$$

- (c). Si dos variables aleatorias son conjuntamente Gaussianas, entonces, cada una de estas también tiene FDP Gaussiana. La afirmación inversa sólo se cumple cuando las variables aleatorias son independientes entre sí.

**Ejemplo 2.13.** Sea la FDP conjunta Gaussiana (2.47) de un fenómeno aleatorio descrito por las variables aleatorias  $(x, y)$  con parámetros  $m_\xi, \sigma_\xi^2$ ,  $\xi \in x, y$  y  $\rho_{xy}$ . Determinar las densidades marginales  $p_\xi(x), p_\xi(y)$ , así como las FDP condicionales  $p_\xi(y|x)$  y  $p_\xi(x|y)$ .

La densidad marginal se halla a partir de la expresión (2.39), en la cual para la FDP conjunta Gaussiana (2.47), se puede realizar el siguiente cambio de variables:

$$(x - m_x) / \sqrt{2}\sigma_x = u, \quad (y - m_y) / \sqrt{2}\sigma_y = v$$

con lo cual se obtiene

$$p_\xi(x) = \frac{1}{\sqrt{2\pi}\sigma_x\sqrt{1-\rho_{xy}^2}} \exp\left(-\frac{u^2}{1-\rho_{xy}^2}\right) \int_{-\infty}^{\infty} \exp\left(-\frac{1}{1-\rho_{xy}^2}v^2 + \frac{2\rho_{xy}u}{1-\rho_{xy}^2}v\right) dv$$

Teniendo en cuenta que

$$\int_{-\infty}^{\infty} \exp(-p^2 x^2 \pm qx) dx = \frac{\sqrt{\pi}}{p} \exp\left(\frac{q^2}{4p^2}\right)$$

se halla la densidad respectiva marginal

$$p_{\xi}(x) = \frac{1}{\sqrt{2\pi}\sigma_x} e^{-u^2} = \frac{1}{\sqrt{2\pi}\sigma_x} \exp\left(-\frac{(x - m_{1x})^2}{2\sigma_x^2}\right)$$

De la expresión anterior, es clara la estructura Gaussiana del variable  $x$  con media  $m_{1x}$  y varianza  $\sigma_x^2$ . De forma similar, se encuentra la densidad marginal alterna por la variable  $y$ , que también resulta ser de estructura Gaussiana con momentos  $m_{1y}, \sigma_y^2$ . Así mismo, de (2.41) se hallan las correspondientes FDP condicionales:

$$p(y|x) = \frac{1}{\sigma_y \sqrt{2\pi} \sqrt{1 - \rho_{xy}^2}} \exp\left(-\frac{1}{2(1 - \rho_{xy}^2)} \left(y - m_{1y} - \rho_{xy} \frac{\sigma_y}{\sigma_x} (x - m_{1x})\right)^2\right)$$

$$p(x|y) = \frac{1}{\sigma_x \sqrt{2\pi} \sqrt{1 - \rho_{xy}^2}} \exp\left(-\frac{1}{2(1 - \rho_{xy}^2)} \left(x - m_{1x} - \rho_{xy} \frac{\sigma_x}{\sigma_y} (y - m_{1y})\right)^2\right)$$

las cuales corresponden a las respectivas variables Gaussianas con parámetros

$$m_{y|x} = m_y + \rho_{xy} \frac{\sigma_y}{\sigma_x} (x - m_{1x}), \sigma_{y|x} = \sigma_y \sqrt{1 - \rho_{xy}^2}$$

$$m_{x|y} = m_x + \rho_{xy} \frac{\sigma_x}{\sigma_y} (y - m_{1y}), \sigma_{x|y} = \sigma_x \sqrt{1 - \rho_{xy}^2}$$

**Transformación de FDP en modelos con dos dimensiones.** Sea el par de variables aleatorias  $\xi_i, i = 1, 2$ , definidas por la relación mutua:  $\eta_i = g_i(\xi_1, \xi_2)$ ,  $i = 1, 2$ , siendo  $g_i, i = 1, 2$  funciones determinísticas conocidas. Por cuanto, la distribución conjunta se da por la relación:

$$F_{\eta_1 \eta_2}(y_1, y_2) = P\{\eta_1 < y_1, \eta_2 < y_2\} = P\{g_1(\xi_1, \xi_2) < y_1, g_2(\xi_1, \xi_2) < y_2\}$$

entonces

$$F_{\eta_1 \eta_2}(y_1, y_2) = \int \int_D p_{\eta_1 \eta_2}(x_1, x_2) dx_1 \quad (2.48)$$

donde el área de integración  $D$  se determina por las desigualdades:

$$g_1(x_1, x_2) < y_1, \quad g_2(x_1, x_2) < y_2$$

La FDP transformada se obtiene diferenciando la expresión (2.48):

$$p_{\eta_1 \eta_2}(y_1, y_2) = \partial^2 F_{\eta_1 \eta_2} / \partial y_1 \partial y_2 \quad (2.49)$$

La densidad obtenida (2.49) se puede calcular directamente, sin necesidad de determinar inicialmente la función  $F_{\eta_1\eta_2}(y_1, y_2)$ , empleando las respectivas expresiones de relación inversa, esto es,  $\xi_i = g_i^{-1}(\eta_1, \eta_2)$ ,  $i = 1, 2$ . Particularmente,

$$p_{\eta_1\eta_2}(y_1, y_2) = p_{\eta_1\eta_2}(g_1^{-1}(y_1, y_2), g_2^{-1}(y_1, y_2)) |J(y_1, y_2)| \quad (2.50)$$

donde

$$J(y_1, y_2) = \frac{\partial(x_1, x_2)}{\partial(y_1, y_2)} = \begin{bmatrix} \partial g_1^{-1}/\partial y_1 & \partial g_1^{-1}/\partial y_2 \\ \partial g_2^{-1}/\partial y_1 & \partial g_2^{-1}/\partial y_2 \end{bmatrix} = \left( \frac{\partial(y_1, y_2)}{\partial(x_1, x_2)} \right)^{-1} \quad (2.51)$$

es el Jacobiano de la transformación de las variables aleatorias  $\{\xi_i\}$  en las variables,  $\{\eta_i\}$ ,  $i = 1, 2$ , respectivamente.

**Ejemplo 2.14.** Sea la FDP conjunta  $p_{\xi_1\xi_2}(x_1, x_2)$  correspondiente a las variables aleatorias  $\xi_1, \xi_2$ . Dados los coeficientes constantes  $a, b, c, d$ , hallar la FDP de 2 dimensiones  $p_{\eta_1\eta_2}(y_1, y_2)$ , para las variables aleatorias  $\eta_1, \eta_2$ , transformadas mediante las relaciones

$$\eta_1 = a\xi_1 + b\xi_2, \quad \eta_2 = c\xi_1 + d\xi_2$$

La solución exige que el determinante del sistema, compuesto por los coeficientes  $a, b, c, d$ , sea diferente de cero. Entonces el sistema compuesto por el par de ecuaciones algebraicas lineales  $y_1 = ax_1 + bx_2$ ,  $y_2 = cx_1 + dx_2$  tiene la única solución:

$$x_1 = a_1y_1 + b_1y_2, \quad x_2 = c_1y_1 + d_1y_2$$

en la cual, los coeficientes  $a_1, b_1, c_1, d_1$  se expresan a través de el respectivo conjunto  $a, b, c, d$ . En este caso, el Jacobiano de la transformación (2.51) tiene la forma:

$$\begin{aligned} J(x_1, x_2) &= \frac{\partial(x_1, x_2)}{\partial(y_1, y_2)} = \left( \frac{\partial(y_1, y_2)}{\partial(x_1, x_2)} \right)^{-1} = \frac{1}{\begin{bmatrix} \partial y_1/\partial x_1 & \partial y_1/\partial x_2 \\ \partial y_2/\partial x_1 & \partial y_2/\partial x_2 \end{bmatrix}} = \frac{1}{\begin{bmatrix} a & b \\ c & d \end{bmatrix}} \\ &= \frac{1}{ad - bc} \end{aligned}$$

Por lo tanto, en correspondencia con (2.50), se obtiene

$$p_{\eta_1\eta_2}(y_1, y_2) = \frac{1}{|ad - bc|} p_{\eta_1\eta_2}(a_1y_1 + b_1y_2, c_1y_1 + d_1y_2)$$

**Ejemplo 2.15.** Sean dos variables Gaussianas independientes  $x$  e  $y$  con valores medios cero,  $m_{1x} = m_{1y} = 0$ , e igual varianza,  $\sigma_x^2 = \sigma_y^2 = \sigma^2$ , que corresponden a la representación cartesiana de un conjunto de puntos. Hallar las densidades de las variables transformadas  $r$  y  $\varphi$ , que representan al módulo y fase de las coordenadas polares y relacionadas con las cartesianas en la forma,  $x = r \cos \varphi$ ,  $y = r \sin \varphi$ .

Debido a la independencia estadística de las variables cartesianas, entonces la FDP conjunta corresponde a la multiplicación de las FDP marginales originales (ver ec. (2.40)),

$$p(x, y) = p(x) p(y) = \frac{1}{2\pi\sigma^2} \exp\left(-\frac{(x^2 + y^2)}{2\sigma^2}\right)$$

La FDP conjunta de dos dimensiones de las variables  $r$  y  $\varphi$ , a partir de la ec. (2.50), son iguales a  $p(r, \varphi) = p(x(r, \varphi), y(r, \varphi)) J(r, \varphi)$ , siendo

$$J(r, \varphi) = \begin{bmatrix} \frac{\partial x}{\partial r} & \frac{\partial x}{\partial \varphi} \\ \frac{\partial y}{\partial r} & \frac{\partial y}{\partial \varphi} \end{bmatrix} = \begin{bmatrix} \cos \varphi & -r \sin \varphi \\ \sin \varphi & r \cos \varphi \end{bmatrix} = r$$

Cada una de la distribuciones marginales de las variables transformadas se halla integrando su FDP conjunta de dos dimensiones por la variable aleatoria alterna,

$$\begin{aligned} p(r) &= \int_{-\pi}^{\pi} p(r, \varphi) d\varphi = \frac{r}{2\pi\sigma^2} \exp\left(-\frac{r^2}{2\sigma^2}\right) \int_{-\pi}^{\pi} d\varphi \\ &= \frac{r}{\sigma^2} \exp\left(-\frac{r^2}{2\sigma^2}\right), \quad r > 0 \end{aligned} \quad (1)$$

La FDP obtenida para el módulo corresponde a la densidad de Rayleigh, cuando  $n = 2$  en la expresión (2.31) para la  $\chi^2$ -densidad. De manera similar, se halla la FDP de la fase,

$$\begin{aligned} p(\varphi) &= \int_0^{\infty} p(r, \varphi) dr = \frac{1}{2\pi} \int_0^{\infty} \frac{r}{\sigma^2} \exp\left(-\frac{r^2}{2\sigma^2}\right) dr \\ &= \frac{1}{2\pi}, \quad -\pi < \varphi < \pi \end{aligned}$$

A partir de las relaciones obtenidas para la transformación para FDP de dos dimensiones, se pueden hallar los momentos conjuntos de las funciones  $\zeta_i = g_i(\xi, \eta)$ ,  $i = 1, 2$ , de manera directa tomando como base la densidad conjunta  $p_{\xi\eta}(x, y)$  de las respectivas variables aleatorias  $\xi$  y  $\eta$ . En este caso la esperanza matemática, la varianza y la función de correlación se determinan de forma correspondiente por las siguientes expresiones:

$$m_{1\zeta_i} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g_i(x, y) p_{\xi\eta}(x, y) dx dy, \quad i = 1, 2 \quad (2.52a)$$

$$\sigma_{\zeta_i}^2 = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (g_i(x, y) - m_{1\zeta_i})^2 p_{\xi\eta}(x, y) dx dy, \quad i = 1, 2 \quad (2.52b)$$

$$R_{\zeta_1\zeta_2} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (g_1(x, y) - m_{1\zeta_1})(g_2(x, y) - m_{1\zeta_2}) p_{\xi\eta}(x, y) dx dy \quad (2.52c)$$

En particular, de (2.52a), (2.52b) y (2.52c) se tienen las siguientes propiedades:

(a). Sea  $c$  una variable no aleatoria, entonces

$$E\{c\} = c, \quad E\{c\xi\} = cE\{\xi\}$$



(b). Para cualquier par de variables aleatorias  $\xi$  y  $\eta$ , se cumple que

$$E\{\xi\eta\} = E\{\xi\}E\{\eta\} - R_{xy}$$

Por cierto, si se cumple que ambas variables  $\xi$  y  $\eta$  son no correlacionadas, esto es, tienen correlación nula,  $R_{xy} = 0$ , entonces,  $E\{\xi\eta\} = E\{\xi\}E\{\eta\}$ .

(c). Sea  $c$  una variable no aleatoria, entonces

$$\sigma_c^2 = 0, \quad \sigma_{c\xi}^2 = c^2\sigma_\xi^2, \quad \sigma_{c\xi} = |c|\sigma_\xi$$

(d). Para cualquier par de variables aleatorias  $\xi$  y  $\eta$ ,

$$\sigma_{\xi \pm \eta}^2 = \sigma_\xi^2 + \sigma_\eta^2 \pm 2R_{xy}(\tau)$$

Si ambas variables aleatorias  $\xi$  y  $\eta$  tienen correlación cero, entonces se cumple que,

$$\sigma_{\xi \pm \eta}^2 = \sigma_\xi^2 + \sigma_\eta^2$$

(e). Para cualquier par de variables aleatorias independientes  $\xi$  y  $\eta$

$$\sigma_{\xi\eta}^2 = \sigma_\xi^2\sigma_\eta^2 + m_{1\xi}^2\sigma_\eta^2 + m_{1\eta}^2\sigma_\xi^2$$

**Ejemplo 2.16.** Hallar la media, la varianza y función de correlación de las variables aleatorias  $\{\zeta_i, i = 1, 2\}$ , expresadas en función de las variables  $\xi$  y  $\eta$ , en la forma aditiva,

$$\zeta_i = k_{1i}\xi + k_{2i}\eta, \quad k_{ni} = \text{const.}, \quad n, i = 1, 2.$$

Basados en las propiedades, expuestas anteriormente, se obtiene

$$\begin{aligned} E\{\zeta_i\} &= E\{k_{1i}\xi + k_{2i}\eta\} = k_{1i}E\{\xi\} + k_{2i}E\{\eta\} \\ \sigma_{\zeta_i}^2 &= \sigma_{k_{1i}\xi + k_{2i}\eta}^2 = k_{1i}^2\sigma_\xi^2 + k_{2i}^2\sigma_\eta^2 + 2R_{xy} \\ &= k_{1i}^2\sigma_\xi^2 + k_{2i}^2\sigma_\eta^2 + 2k_{1i}k_{2i}\sigma_\xi\sigma_\eta\rho_{\xi\eta} \\ R_{\zeta_1\zeta_2} &= E\{(\zeta_1 - m_{1\zeta_1})(\zeta_2 - m_{1\zeta_2})\} = E\{\zeta_1\zeta_2\} - E\{\zeta_1\}E\{\zeta_2\} \\ &= E\{(k_{11}\xi + k_{21}\eta)(k_{12}\xi + k_{22}\eta)\} - (k_{11}E\{\xi\} + k_{21}E\{\eta\})(k_{12}E\{\xi\} + k_{22}E\{\eta\}) \\ &= k_{11}k_{12}\sigma_\xi^2 + k_{21}k_{22}\sigma_\eta^2 + (k_{21}k_{12} + k_{11}k_{22})R_{\xi\eta} \\ &= k_{11}k_{12}\sigma_\xi^2 + k_{21}k_{22}\sigma_\eta^2 + \sigma_\xi\sigma_\eta\rho_{\xi\eta} \end{aligned}$$

**Modelos de probabilidad de  $n$  dimensiones.** Los diversos modelos de FDP se generalizan para el caso de un número finito de dimensiones  $n$ , correspondiente a las variables aleatorias  $x_1, \dots, x_n$  del proceso  $\xi$ . En este caso, la generalización de (2.38) para la distribución con dimensión  $F_\xi(x_1, \dots, x_n)$  corresponde a la probabilidad de que conjuntamente

cada variable aleatoria  $x_k$  tome el respectivo valor menor que  $\lambda_k$ ,  $k = 1, \dots, n$ , esto es,

$$\begin{aligned} P\{x_1 \leq \lambda_1, x_2 \leq \lambda_2, \dots, x_n \leq \lambda_n\} &= F_\xi(\lambda_1, \lambda_2, \dots, \lambda_n) \\ &= \int_{-\infty}^{\lambda_1} \int_{-\infty}^{\lambda_2} \cdots \int_{-\infty}^{\lambda_n} p_\xi(x_1, x_2, \dots, x_n) dx_1 dx_2 \cdots dx_n \end{aligned}$$

La generalización de las propiedades obtenidas para las FDP de 2 dimensiones, extendida al caso de variables aleatorias con mayor dimensión, corresponde a las expresiones:

(a).  $0 \leq F_\xi(\lambda_1, \lambda_2, \dots, \lambda_n) \leq 1$ . Donde

$$\begin{aligned} F_\xi(-\infty, x_2, \dots, x_n) &= F_\xi(x_1, -\infty, \dots, x_n) = F_\xi(x_1, x_2, \dots, -\infty) = 0 \\ F_\xi(\infty, \dots, \infty, x_{m+1}, \dots, x_n) &= F_\xi(x_{m+1}, \dots, x_n), \quad \forall m \leq n \end{aligned}$$

(b).  $\frac{\partial}{\partial x_i} F_\xi(\lambda_1, \lambda_2, \dots, \lambda_n) \geq 0$ ,  $i = 1, 2, \dots, n$ . De lo cual, resulta que

$$F_\xi(x_1, \dots, \lambda_{k1}, \dots, x_n) \geq F_\xi(x_1, \dots, \lambda_{k2}, \dots, x_n), \quad \forall \lambda_{k1} \geq \lambda_{k2}, \lambda_{k1}, \lambda_{k2} \in x_k$$

(c). Si existe la respectiva función densidad de probabilidad, entonces

$$p_\xi(x_1, x_2, \dots, x_n) = \frac{\partial^n F_\xi(x_1, x_2, \dots, x_n)}{\partial x_1 \partial x_2 \cdots \partial x_n} \geq 0$$

Además, se cumple que,

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} p_\xi(x_1, x_2, \dots, x_n) dx_1 dx_2 \cdots dx_n = 1$$

Las variables aleatorias con dimensión  $n$  se denominan *mutuamente no correlacionadas* si para las respectivas funciones de probabilidad se cumple que

$$F_\xi(x_1, \dots, x_n) = \prod_{k=1}^n F_\xi(x_k) \quad (2.53a)$$

$$p_\xi(x_1, x_2, \dots, x_n) = \prod_{k=1}^n p_\xi(x_k) \quad (2.53b)$$

Las expresiones (2.53a) y (2.53b) generalizan la relación (2.40). No obstante, aunque se puede afirmar que de un conjunto de variables aleatorias no correlacionadas con dimensión  $n$  resulta la independencia estadística por todas las posibles parejas de variables aleatorias  $x_i$  y  $x_j$ ,  $i, j = 1, \dots, n$ , la afirmación inversa, en forma general, no es cierta [12], excepto para la FDP Gaussiana [15].

El conjunto de variables aleatorias  $\{x_i : i = 1, \dots, n\}$  se puede entender como las coordenadas puntuales o componentes del vector aleatorio  $\boldsymbol{\xi}$  sobre un espacio con dimensión  $n$ . En este caso, para cumplir la condición de independencia estadística de cualquier par de vectores aleatorios  $\{x_i : i = 1, \dots, n\} \in \boldsymbol{\xi}$  y  $\{y_i : i = 1, \dots, n\} \in \boldsymbol{\eta}$  es necesario y suficiente que la densidad conjunta de sus componentes sea igual al producto de las FDP de las componentes de cada uno de los vectores:

$$p(\boldsymbol{\xi}, \boldsymbol{\eta}) = p_{\boldsymbol{\xi}\boldsymbol{\eta}}(x_1, \dots, x_n, y_1, \dots, y_n) = p_{\boldsymbol{\xi}}(x_1, \dots, x_n) p_{\boldsymbol{\eta}}(y_1, \dots, y_n)$$

El valor medio de la variable  $x_k$ , para el caso de modelos de FDP con dimensión múltiple se generaliza en la forma,

$$m_{1\xi}(x_k) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} x_k p_{\xi}(x_1, \dots, x_n) dx_1 \dots dx_n$$

La respectiva covarianza del par de variables  $x_k$  y  $x_l$ ,  $\forall k, l = 1, \dots, n$ , es igual a,

$$\mu_{11}(x_k, x_l) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} (x_k - m_{1\xi}(x_k))(x_l - m_{1\xi}(x_l)) p_{\xi}(x_1, \dots, x_n) dx_1 \dots dx_n,$$

Cabe anotar, que para el caso de  $k = l$ , el momento  $\mu_{11}(x_k, x_k)$  coincide con la varianza de  $x_k$ , esto es,  $\mu_{11}(x_k, x_k) = \sigma_{x_k}^2$ .

En forma general, los momentos centralizados conjuntos de cualquier orden se determinan por la expresión

$$\begin{aligned} \mu_{k_1 k_2 \dots k_n}(x_1, \dots, x_n) \\ = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} (x_1 - m_{1\xi}(x_1))^{k_1} \cdots (x_n - m_{1\xi}(x_n))^{k_n} p_{\xi}(x_1, \dots, x_n) dx_1 \dots dx_n \end{aligned}$$

siendo  $k_l \in \mathbb{N}$ ,  $l = 1, \dots, n$ .

Usualmente, cuando se tiene una cantidad de variables  $n > 2$ , es preferible emplear la *matriz de covarianza*, definida como

$$\mathbf{K}_{n \times n} \triangleq \begin{bmatrix} K_{x_1 x_1} & K_{x_1 x_2} & \cdots & K_{x_1 x_n} \\ K_{x_2 x_1} & K_{x_2 x_2} & \cdots & K_{x_2 x_n} \\ \vdots & \vdots & \ddots & \vdots \\ K_{x_n x_1} & K_{x_n x_2} & \cdots & K_{x_n x_n} \end{bmatrix} \quad (2.54)$$

De la definición (2.44) para la función de covarianza resulta que  $K_{x_k x_l} = K_{x_l x_k}$ , además

$K_{x_k x_k} = \sigma_{x_k}^2$ , por lo tanto es frecuente el empleo de la siguiente matriz triangular

$$\mathbf{K} = \begin{bmatrix} \sigma_{x_1}^2 & K_{x_1 x_2} & \cdots & K_{x_1 x_n} \\ & \sigma_{x_2}^2 & \cdots & K_{x_2 x_n} \\ & & \cdots & \cdots \\ & & & \sigma_{x_n}^2 \end{bmatrix}$$

Si el conjunto de variables  $\{x_k : 1, \dots, n\}$  cumple las condiciones de independencia estadística, entonces los momentos cruzados centrales de primer orden son iguales a cero, con lo cual se obtiene la siguiente matriz diagonal:

$$\mathbf{K} = \begin{bmatrix} \sigma_{x_1}^2 & 0 & 0 & 0 \\ & \sigma_{x_2}^2 & \cdots & 0 \\ & & \cdots & \cdots \\ & & & \sigma_{x_n}^2 \end{bmatrix}$$

**Modelos Gaussianos con  $n$  dimensiones.** Los modelos de distribución Gaussiana conjunta con dimensión  $n > 2$ , es preferible, expresarlos en forma matricial:

$$p_{\xi}(\mathbf{x}) = \frac{1}{(2\pi)^{n/2} |\mathbf{K}|^{1/2}} \exp \left( -\frac{1}{2} (\mathbf{x} - \mathbf{m}_{1\mathbf{x}})^{\top} \mathbf{K}^{-1} (\mathbf{x} - \mathbf{m}_{1\mathbf{x}}) \right) \quad (2.55)$$

siendo  $\mathbf{x}$  y  $\mathbf{m}$  los vectores columna conformados por los elementos  $\{x_k : k = 1, \dots, n\}$  y  $\{m_{1x_k} : k = 1, \dots, n\}$ , respectivamente. Además,  $|\mathbf{K}| = \det(\mathbf{K})$ .

De la expresión (2.55) se observa, que cuando el conjunto de variables aleatorias Gaussianas  $\{x_k\}$  tiene covarianza nula, esto es,  $\{K_{x_k x_l} = 0 : \forall k \neq l; k, l = 1, \dots, n\}$ , entonces, las variables cumplen la condición de independencia estadística dada en (2.53b).

En forma general, se puede demostrar que las siguientes son las respectivas relaciones recursivas en los momentos iniciales y centralizados para la FDP Gaussiana :

$$m_{k\xi} = \frac{k(k-1)}{2} \int_0^{\sigma_{\xi}^2} m_{k-2,\xi} (m_{1\xi}, \sigma_{\xi}^2) d\sigma_{\xi}^2 + m_{1\xi}^k \quad (2.56a)$$

$$\mu_k = \frac{k(k-1)}{2} \int_0^{\sigma_{\xi}^2} \mu_{k-2} d\sigma_{\xi}^2 = \begin{cases} 1 \times 3 \times 5 \dots (k-1) \sigma_{\xi}^k, & k \in \text{par} \\ 0, & k \in \text{impar} \end{cases} \quad (2.56b)$$

### 2.1.5. Medidas de información en variables aleatorias

**Información en señales discretas.** La medida de información contenida en un proceso aleatorio, que toma los valores discretos  $\{x_n : n = 1, \dots, N\}$ , cumple las restricciones:

1. La cantidad de información debe ser un valor aditivo, esto es, en dos procesos aleatorios independientes la medida es igual a la suma de las cantidades en cada proceso.

2. La cantidad de información en un proceso no aleatorio (determinístico) es cero.
3. La cantidad de información no debe depender de algún factor subjetivo.

De lo anterior, la información propia de un evento  $x_i$ , también denominado *símbolo*, se asume mediante la medida,  $I(x_i) = -\log_2 p(x_i)$ , donde  $p(x_i)$  es la probabilidad a priori de aparición del valor  $x_i$ . El logaritmo se toma en base 2 y, en consecuencia, las unidades de información son *bits* (en adelante el índice de la base se omitirá).

Otras medidas de información se determinan por las siguientes relaciones:

- Información propia condicional de la variable  $x_m$ , conocida la variable aleatoria  $y_n$ ,

$$I(x_m|y_n) = -\log p(x_m|y_n)$$

- Información mutua de dos variables o información de  $y_n$  con respecto a  $x_m$ ,

$$I(x_m; y_n) = \log p(x_m|y_n) \quad (2.57)$$

- Información propia del evento conjunto  $(x_m, y_n)$ ,

$$I(x_m, y_n) = -\log p(x_m, y_n)$$

- Cantidad media de información, presente en  $x_n$  como elemento del alfabeto  $\{x_n\}$ ,

$$I(X; y_m) = \sum_{n=1}^N I(x_n; y_m) p(x_n|y_m) = \sum_{n=1}^N p(x_n|y_m) \log \frac{p(x_n|y_m)}{p(x_n)}$$

- Cantidad media de información mutua sobre los elementos correspondientes a un alfabeto  $Y = \{y_m : m = 1 \dots, M\}$  para un valor determinado de  $x_n$ ,

$$I(x_n; Y) = \sum_{m=1}^M I(x_n; y_m) p(y_m|x_n) = \sum_{m=1}^M p(y_m|x_n) \log \frac{p(y_m|x_n)}{p(y_m)}$$

- Valor medio de la cantidad completa de información mutua entre conjuntos  $(Y, X)$ ,

$$I(X; Y) = \sum_{n=1}^N \sum_{m=1}^M p(x_n, y_m) \log \frac{p(x_n|y_m)}{p(x_n)} = \sum_{m=1}^M p(y_m) I(X; y_m)$$

Si  $x_n$  se relaciona estadísticamente con los valores  $y_m$  y  $z_k$ ,  $k = 1, \dots, K$ , conocidas las FDP condicionales  $p(x_n, y_m, z_k)$ , la información condicional mutua se determina por,

$$I(x_m; y_n|z_k) \log \frac{p(x_m|y_n, z_k)}{p(x_m|z_k)} = \log \frac{p(x_m; y_n|z_k)}{p(x_m|z_k) p(y_n|z_k)}$$

Las siguientes relaciones son válidas para las cantidades puntuales de información:

$$\begin{aligned}
I(x_n; y_m) &= I(y_m; x_n), I(x_n; y_m) \leq I(x_n), I(x_n; y_m) \leq I(y_m) \\
I(x_n; y_m, z_k) &= I(x_n; y_m) + I(x_n; z_k | y_m) = I(x_n; z_k) + I(x_n; y_m | z_k) \\
I(x_n; y_m) &= I(x_n) - I(x_n | y_m) = I(y_m) - I(y_m | x_n) = I(x_n) + I(y_m) - I(x_n, y_m) \\
I(x_n, y_m) &= I(x_n) + I(y_m) - I(x_n; y_m) \\
I(X; y_m) &\geq 0, I(x_n, Y) \geq 0 \\
I(X, Y) &= I(Y, X) \geq 0, \\
I(X; Y, Z) &= I(X, Y) + I(X; Z | Y) \\
I(Y, Z; X) &= I(Y; X) + I(Z; X | Y)
\end{aligned}$$

De (2.8), se define el valor medio de la información propia para un alfabeto discreto:

$$I(X) = E\{I(x_n)\} = \sum_{n=1}^N p(x_n) I(x_n) = - \sum_{n=1}^N p(x_n) \log p(x_n) \triangleq H(X) \quad (2.58)$$

El valor  $H(X)$  se denomina *entropía* de la variable aleatoria  $X$  y corresponde a la medida cuantitativa de su incertidumbre. La entropía es la principal característica de un fuente de información: entre más alto sea el valor de la entropía mayor es la información contenida por el conjunto  $X$ .

La entropía tiene las siguientes propiedades:

- (a). *Definida positiva:*  $H(X) \geq 0$ ,
- (b). *Aditividad.* Sea  $H(Y|X)$  la entropía condicional de los valores del conjunto  $Y$ , dado un conjunto de eventos  $X$ , entonces, la entropía  $H(Y, X)$  del conjunto de eventos conjuntos  $X$  y  $Y$  se define como:

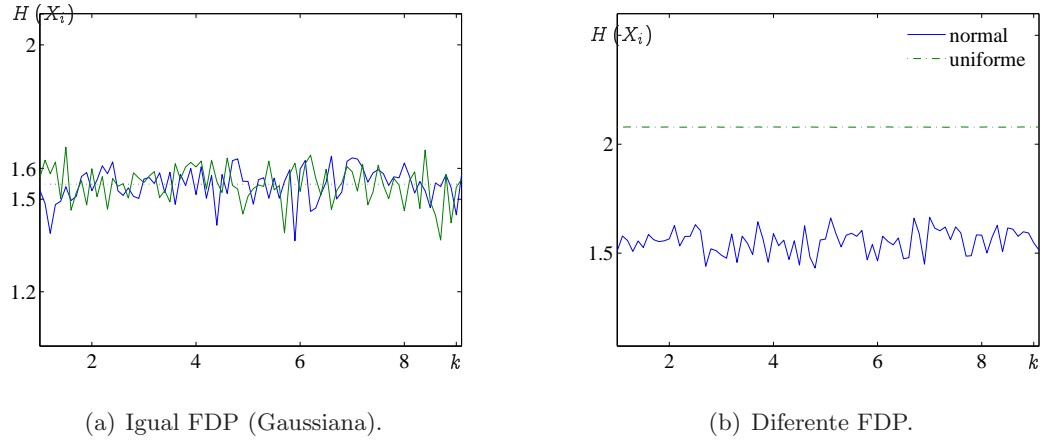
$$H(Y, X) = H(X) + H(Y|X) = H(Y) + H(X|Y) \quad (2.59)$$

Por cierto, si los conjuntos  $X$  e  $Y$  son independientes, entonces  $H(Y|X) = H(Y)$ , por lo tanto, se cumple que  $H(Y, X) = H(X) + H(Y)$ .

- (c). *Valor acotado.* Si el conjunto de valores  $X = \{x_n : n = 1, \dots, N\}$ , la máxima entropía contenida está acotada por el valor  $H(X) \leq \lg N$ , donde la igualdad tiene lugar cuando todos los valores  $x_n$  son equiprobables e independientes estadísticamente.

**Ejemplo 2.17.** Desarrollar un programa que simule dos observaciones con longitud  $N$ , para las variables aleatorias  $X_i \in \{x_l : l = 1, \dots, L\}$ ,  $i = 1, 2$ , que permita analizar el comportamiento de sus respectivos valores de entropía para los siguientes casos:

- Ambas variables tienen FDP Gaussiana, con valor medio cero. Estimar las entropías para valores de relación de las varianzas:  $\sigma_{X_1}^2 = k\sigma_{X_2}^2$ ,  $k = 1, \dots, 10$ .
- La variable  $X_1$  tiene FDP Gaussiana, mientras la variable  $X_2$  se genera con FDP uniforme; ambas con los mismos valores medio y de varianza.



**Figura 2.9.** Entropía en función de  $k = \sigma_{X_2}^2 / \sigma_{X_1}^2$ ,  $N = 4096$  y  $L = 8$ .

En la Figura 2.9(a), se muestran los resultados para el caso de generación de ambas variables Gaussianas con media  $m_{1X_1} = m_{1X_2} = 0$ , pero  $\sigma_{X_1}^2 = k\sigma_{X_2}^2$ . Se observa que la estimación de la entropía converge a un mismo valor para ambos casos de análisis. La Figura 2.9(b) corresponde al caso de diferentes FDP (Gaussiana y uniforme) y muestra que el valor de la entropía, para dos sucesiones aleatorias con iguales valores medio y de varianza, depende de la estructura de aleatoriedad.

**Ejemplo 2.18.** Demostrar que para dos sucesos con probabilidades respectivas,  $p_1$  y  $p_2$ , la mayor entropía se obtiene para  $p_1 = p_2$ .

Al reemplazar en la definición de entropía (2.58),  $p_2 = 1 - p_1$ , se tiene que

$$H = -(p_1 \ln p_1 + p_2 \ln p_2) = -(p_1 \ln p_1 + (1 - p_1) \ln (1 - p_1))$$

El máximo valor de  $p_1$  se obtiene al diferenciar e igualar a cero la anterior ecuación,

$$\frac{dH}{dp_1} = - \left( p_1 \frac{1}{p_1} + \ln p_1 - \ln (1 - p_1) \right) = 0$$

Como resultado, al resolver la ecuación se obtiene que  $p_1 = 0.5$ , luego,  $p_1 = p_2$ .

La entropía condicional se define como:

$$H(Y|X) = - \sum_{n=1}^N \sum_{m=1}^M p(x_n, y_m) \log p(y_m|x_n) = -p(x_n) \sum_{m=1}^M p(y_m|x_n) \log p(y_m|x_n)$$

para la cual se cumple que,  $0 \leq H(Y|X) \leq H(Y)$ .

El valor medio de la información mutua  $I(X; Y)$  se relaciona con la entropía como:

$$\begin{aligned} I(X; Y) &= H(X) - H(X|Y) = H(Y) - H(Y|X) = H(X) + H(Y) - H(X, Y) \\ I(X; Y) &\leq H(X), \quad I(X; Y) \leq H(Y) \end{aligned}$$

La entropía es una medida adecuada de la incertidumbre contenida en la FDP y su uso es preferible en aquellos casos cuando la densidad es asimétrica, con muchos picos de concentración, que es cuando los momentos pierden su efecto interpretativo del proceso.

De otra parte, la medida de incertidumbre  $H(X)$  depende de la relación estadística entre los elementos  $\{x_i \in X\}$ , por lo que se puede definir qué tanto de la entropía máxima potencial no se contiene en el conjunto  $X$ , a través de la medida de relación denominada *redundancia*:

$$R = \frac{H_{\max} - H(X)}{H_{\max}} = \frac{\log N - H(X)}{\log N}, \quad 0 \leq R \leq 1 \quad (2.60)$$

**Información en señales continuas.** Por cuanto la cantidad de valores, que se contienen en cualquier rango continuo, es infinita, así sea dado sobre un intervalo finito, entonces la cantidad de información en una variable continua también es infinita, así como su entropía. Por lo tanto, se define la *entropía diferencial* de la variable aleatoria continua  $x$ :

$$h(x) = E \left\{ \log \frac{1}{p(x)} \right\} = \int_X p(x) \log \frac{1}{p(x)} dx \quad (2.61)$$

la cual no se puede analizar como una medida de información propia, a diferencia de lo que ocurre con la entropía de las variables discretas. La medida  $h(x)$  no tiene las mismas propiedades que la entropía de las variables discretas, en particular la entropía diferencial puede tomar valores negativos. El sentido informativo no está en la misma entropía diferencial  $h(x)$ , sino en la diferencia de dos valores de  $h(x)$ , motivo por el cual se da el nombre de entropía diferencial.

**Ejemplo 2.19.** Hallar la entropía diferencial para los casos de FDP uniforme (2.33) y Gaussiana (2.23).

En correspondencia con (2.61), se tiene

– *FDP uniforme.*

$$h_U(x) = - \int_a^b p(x) \ln p(x) dx = - \frac{1}{b-a} \ln \frac{1}{b-a} \int_a^b dx = \ln(b-a)$$

Por cuanto se puede demostrar que  $b-a = 2\sqrt{3}\sigma_x$ , entonces,  $h_U(x) = \ln(2\sqrt{3}a)$

– *FDP Gaussiana.*

$$\begin{aligned} h_N(x) &= - \int_a^b \mathcal{U}(m_{1x}, \sigma_x) \ln \mathcal{U}(m_{1x}, \sigma_x) dx = \\ &= \ln(\sqrt{2\pi e}\sigma) \end{aligned}$$

Como se observa de las relaciones obtenidas, la diferencia de valores entre las ambas



entropías corresponde a

$$\Delta h = h_{\mathcal{N}}(x) - h_{\mathcal{U}}(x) = \ln \sqrt{2\pi e} \sigma_{\mathcal{N}(x)} - \ln (2\sqrt{3} \sigma_{\mathcal{U}(x)})$$

Si se asumen las varianzas de ambas distribuciones iguales  $\sigma_{\mathcal{U}(x)}^2 = \sigma_{\mathcal{U}(x)}^2 = \sigma_x^2$ , entonces el valor final de diferencia se obtiene como

$$\Delta h = \ln \frac{\sqrt{2\pi e}}{2\sqrt{3}} = 0.17$$

De los resultados obtenidos se observa que la entropía diferencial de una variable aleatoria continua, con función densidad de probabilidad Gaussiana o uniforme, no depende del valor medio de la variable aleatoria (por cuanto, un valor constante no lleva información) y solo depende en forma proporcional de la varianza de los valores aleatorios.

---

La propiedad de aditividad se conserva en el caso de la entropía diferencial. La entropía diferencial condicional de las variables  $\mathbf{x}$  e  $\mathbf{y}$ , definida como

$$h(\mathbf{x}|\mathbf{y}) = E \left\{ \log \frac{1}{p(\mathbf{x}|\mathbf{y})} \right\}$$

tiene las mismas propiedades que la entropía condicional de las variables discretas.

Así mismo, de todas las posibles FDP de la variable aleatoria  $\mathbf{x}$ , que tengan un valor fijo de potencia  $\sigma_x^2$ , la mayor entropía diferencial posible ocurre para la densidad de probabilidad Gaussiana.

## Problemas

**Problema 2.1.** Hallar el exceso  $\gamma_{2\xi}$  para las FDP: uniforme, Poisson, Binomial y Gaussiana.

**Problema 2.2.** Sea la variable aleatoria  $\eta$  una función lineal de  $\xi$ :  $\eta = \alpha\xi + \beta$ , donde  $\alpha$  y  $\beta$  son constantes. Encontrar la FDP de la variable  $\eta$  para el caso en que la densidad  $p(\xi)$  sea Gaussiana.

**Problema 2.3.** La variable  $\xi$  con FDP uniforme,  $p(\xi) = 1/(\alpha + \beta)$ ,  $\alpha < \xi < \beta$ ,  $\alpha < \beta$ , pasa por un detector cuadrático  $\eta = \xi^2$ . Determinar y representar:

- (a). La función de distribución  $F(\xi)$ .
- (b). La función de distribución y la función densidad de probabilidad de  $\eta$ .

**Problema 2.4.** La variable aleatoria  $\xi$  se describe por la FDP binomial. Hallar los valores de la media y la varianza de la magnitud  $\eta = \exp(\alpha\xi)$ , para los casos:  $p \neq q$  y  $p = q$ .

**Problema 2.5.** Sea la variable aleatoria  $\xi$  con FDP  $p(\xi) = \alpha \exp(-\beta|x|)$ ,  $-\infty < x < \infty$ , donde  $\alpha$  y  $\beta$  son constantes. Encontrar la relación que deben cumplir las constantes.

- (a). Calcular la función de distribución de la magnitud  $\xi$ .
- (b). Dibujar la función de distribución y la FDP para  $\beta = 2$ .

**Problema 2.6.** Sea la fase aleatoria  $\Phi$  uniformemente distribuida en el intervalo  $(0, 2\pi)$ . Calcular la FDP de la variable aleatoria de amplitud  $\xi(t) = A \cos^n(\omega t + \Phi)$ , donde  $\omega$  y  $t$  son constantes. Considérese  $n = 1$ ,  $n = 2$ .

**Problema 2.7.** La variable aleatoria  $\xi$  tiene FDP de Laplace,  $p(\xi) = 0.5t \exp(-t|\xi|)$ ,  $t > 0$ . Hallar la media y la varianza de la variable aleatoria.

**Problema 2.8.** Hallar  $\xi$  y  $\sigma_\xi^2$  de la variable entera aleatoria distribuida entre  $[0, N]$ , para la cual  $\sum_k^N p(\xi_k) = 1$ .

**Problema 2.9.** Hallar los momentos iniciales de la FDP de Weibull, definida en (2.29). Analizar los casos particulares de  $c = 1$  (FDP exponencial).

**Problema 2.10.** Demostrar que para la FDP de Cauchy no existe ningún momento inicial.

**Problema 2.11.** Sea  $\xi$  una variable aleatoria, para la cual converge el segundo momento inicial, asumiendo que  $c$  es una constante. Hallar para qué valor de  $c$  se obtiene el mínimo valor de  $m_{1\eta}$ , donde  $\eta = \xi - c$ .

## 2.2. Estimación en variables aleatorias

Hasta ahora, las definiciones, dadas para los diferentes valores y momentos de aleatoriedad, han supuesto que la cantidad de información, dispuesta sobre cualquier señal aleatoria, es completa. Debido a restricciones de costos, tiempo y posibilidad de tomar toda la información sobre los objetos pertenecientes a la señal aleatoria en análisis, en la realidad se tiene alguna trayectoria  $\mathbf{x}$  o conjunto de segmentos  $\{x_i\}$  tomados de la variable observada  $\mathbf{x} \in \xi$ . Por esto, es necesario calcular de manera aproximada cada valor de aleatoriedad a partir de la trayectoria disponible. En particular, a partir del análisis mediante una función  $g_\theta(x_i)$  sobre los resultados experimentales de una observación con longitud  $n$ ,  $\{x_i : i = 1, \dots, n\}$  se obtiene la característica estadística, o *estimación*,  $\tilde{\theta}$  del valor aleatorio  $\theta$ , entonces:

$$\tilde{\theta} = g_\theta(x_1, \dots, x_n) \quad (2.62)$$

Es necesario considerar, que al repetir el proceso de estimación, debido a la aleatoriedad que presentan los valores en cada  $k$  observación  $\{x_{lk} : l = 1, \dots, n\}$ , entonces la estimación  $\tilde{\theta} = g_\theta(x_{1k}, \dots, x_{nk})$  también es otra variable aleatoria.

Las características probabilísticas de  $\tilde{\theta}$  dependen de factores tales como, la estructura de aleatoriedad de cada variable  $\{x \in \xi : x_1, \dots, x_n\}$ , de la longitud de la observación  $n$  y del tipo de función  $g_\theta(x_i)$  empleada para el análisis de los datos. En este sentido, usualmente, la precisión en la aproximación del cálculo de la característica se describe por la *potencia del error* definida en la forma:

$$\varepsilon^2 = E\{(\tilde{\theta} - \theta)^2\} \quad (2.63)$$

que también corresponde a una variable aleatoria, motivo por el cual, a la función  $g_\theta(x_i)$  se le imponen las siguientes restricciones propias de todas las estimaciones:

1. *Ausencia de sesgo.* Caracteriza el comportamiento de la estimación cuando se amplía la longitud de la observación,  $n \rightarrow \infty$ :

$$E\{\tilde{\theta}\} = \lim_{N \rightarrow \infty} E\{g_\theta(x_1, \dots, x_n)\} = \theta$$

2. *Consistencia.* Analiza la convergencia en el sentido probabilístico del valor de la estimación al aumentar la longitud de la observación:

$$\lim_{n \rightarrow \infty} P\{|\tilde{\theta} - \theta| < \varepsilon\} = P\{|g_\theta(x_1, \dots, x_n) - \theta| < \varepsilon\} = 1$$

Basados en la desigualdad de Chebyshev (2.12), se puede demostrar que la condición suficiente para la consistencia de la estimación es [15]:

$$\lim_{n \rightarrow \infty} E\{(\tilde{\theta} - \theta)^2\} = 0$$

3. *Efectividad.* Cuando se tienen longitudes limitadas de la observación, los diversos

métodos de estimación conocidos pueden presentar diferentes valores de potencia de error (2.63). Entonces, se considera que entre menor sea el valor obtenido del error, para el correspondiente método de estimación, los valores calculados estarán agrupados en distancias más cercanas alrededor del valor  $\theta$ :

$$\epsilon^2 = \min_{\tilde{\theta}} E \{ (\tilde{\theta} - \theta)^2 \} \quad (2.64)$$

4. *Suficiencia.* Cuando la función densidad de probabilidad de la observación  $p(\theta; \tilde{\theta})$  se puede representar en la forma:

$$p(\theta; x_1, \dots, x_n) = p(\theta; \tilde{\theta}) h(x_1, \dots, x_n) = p(\theta; g_\theta(x_1, \dots, x_n)) h(x_1, \dots, x_n)$$

donde  $h(x_1, \dots, x_n)$  no depende de  $\theta$ . Esto es, toda la información contenida en la observación con relación al parámetro  $\theta$ , también está contenida en  $\tilde{\theta}$ .

En la práctica es usual el empleo del error relativo de estimación:

$$\Delta \tilde{\theta} = |\tilde{\theta} - \theta| \quad (2.65)$$

con lo que las primeras tres propiedades de la estimación se expresan en la respectiva forma:

- (a).  $E\{\Delta \tilde{\theta}\} = 0$
- (b).  $\lim_{n \rightarrow \infty} \sigma^2(\Delta \tilde{\theta}) = 0$
- (c).  $\sigma^2(\Delta \tilde{\theta}_k) = \min_{\tilde{\theta} \forall k} \sigma^2(\Delta \tilde{\theta})$

Debido a la naturaleza aleatoria de la estimación y, en particular de su potencia de error (2.63), es necesario determinar un intervalo de valores (y no solamente un valor), en el cual con un nivel de confianza, a priori dado, se localice el valor de la estimación. De otra parte, el intervalo con longitud  $(\theta_{\min}, \theta_{\max})$ , cuyo centro sirve de estimación de una característica aleatoria, el cual con una probabilidad de  $\alpha$  está contenido el valor verdadero de la característica, se denomina *intervalo de confianza* con valor de *significación*  $\alpha$ . En concordancia con lo anterior, existen dos clases de estimaciones: las *puntuales*, cuando el resultado de la aproximación corresponde a un valor y las estimaciones *de intervalo* cuando la aproximación se da dentro de un rango de valores.

### 2.2.1. Estimación puntual de momentos

La suposición, aunque sea con suficientes argumentos, sobre un tipo concreto de FDP,  $p_\xi(x)$ , aún no significa que se tengan los valores concretos de sus respectivos momentos, con lo cual, hay necesidad de estimarlos sobre los valores de la observación  $x_1, \dots, x_n$ .

Existen diversos métodos para la construcción de estimadores puntuales, sobre una observación dada, entre los principales están los siguientes:

**Métodos de los momentos.** Se basa en la equivalencia asumida de los momentos de estimación con los momentos de la FDP teórica, obteniéndose el sistema de ecuaciones:

$$m_{k\theta} \equiv \tilde{m}_{k\xi}; \quad k = 1, 2, \dots, d$$

en función de la cantidad  $d$  de las variables desconocidas,  $\theta_1, \theta_2, \dots, \theta_d$ .

En caso de existir una solución única,  $\tilde{\theta} = g_\theta(x_{1k}, \dots, x_{nk})$ , para el sistema de ecuaciones y cuando la función  $g_\theta$  es continua, entonces la estimación obtenida  $\tilde{\theta}_k$  es consistente. En particular, se puede tomar la siguiente equivalencia de momentos:

$$m_{k\theta} = \int_{x \in \xi} \xi^k p_\theta(d\xi) \cong \frac{1}{n} \sum_{l=1}^n x_l^k, \quad k = 1, 2, \dots, d \quad (2.66)$$

---

**Ejemplo 2.20.** Dada la observación  $\{x_k : 1, \dots, n\}$ , analizar las propiedades de estimación del método de momentos (2.66) para el caso de la media y la varianza.

La esperanza matemática del error (2.65) de estimación para un momento inicial de  $k$  orden se determina como:

$$\begin{aligned} \Delta \tilde{m}_{k\xi} &= \tilde{m}_{k\xi} - m_{k\theta} \\ E\{\Delta \tilde{m}_{k\xi}\} &= E\{\tilde{m}_{k\xi} - m_{k\theta}\} = E\left\{\frac{1}{n} \sum_{l=1}^n x_l^k\right\} - m_{k\theta} = \frac{1}{n} \sum_{l=1}^n E\{x_l^k\} - m_{k\theta} \\ &= \frac{1}{n} n m_{k\theta} - m_{k\theta} = 0 \end{aligned}$$

De lo anterior se deduce que la estimación  $\tilde{m}_{k\theta}$  no tiene sesgo.

De manera similar, se halla la varianza del error (2.65) de estimación:

$$\begin{aligned} \sigma^2(\Delta \tilde{m}_{k\xi}) &= \sigma^2(\tilde{m}_{k\xi}) = \sigma^2\left(\frac{1}{n} \sum_{l=1}^n x_l^k\right) = \frac{1}{n^2} n \sigma^2(x^k) = \frac{1}{n} E\{(x^k - m_{k\theta})^2\} \\ &= \frac{1}{n} E\{x^{2k} - 2m_{k\theta} x^k + m_{k\theta}^2\} = \frac{1}{n} E\{x^{2k} - m_{k\theta}^2\} = \frac{1}{n} (m_{2k\xi} - m_{k\theta}^2) \end{aligned}$$

Al hacer  $n \rightarrow \infty$  se tiene  $\sigma^2(\Delta \tilde{m}_{k\xi}) = 0$ . Por lo tanto, la estimación es consistente.

Haciendo  $k = 1$  se obtiene la estimación de la media, para la cual se tiene que

$$E\{\tilde{m}_{1\xi}\} = m_{1\xi}, \quad \sigma^2(\tilde{m}_{1\xi}) = \frac{m_{2\xi} - m_{1\xi}^2}{n} = \frac{\sigma_\xi^2}{n}$$

La estimación de los momentos centralizados se obtiene como:

$$\tilde{\mu}_{k\xi} = \frac{1}{n} \sum_{l=1}^n (x_l - \tilde{m}_{1\xi})^k$$

con error absoluto de estimación:  $\Delta \tilde{\mu}_{k\xi} = \tilde{\mu}_{k\xi} - \mu_{k\theta}$ . Se puede demostrar que para cualquier  $k$ ,  $E\{\Delta \tilde{\mu}_{k\xi}\} \neq 0$ . Sin embargo, de acuerdo a la última expresión obtenida para la varianza de estimación, se tienen los valores asintóticos:  $E\{\Delta \tilde{\mu}_{k\xi}\} \rightarrow 0, \sigma^2(\Delta \tilde{\mu}_{k\xi}) \rightarrow 0, \forall n \rightarrow \infty$ .

---

En general, las estimaciones obtenidas mediante el método de los momentos, (2.66), no son efectivas [17].

**Método de máxima verosimilitud.** Sea  $x \in \xi(\theta)$  una variable aleatoria con FDP que depende únicamente del parámetro  $\theta$ . Cuando se tiene la observación compuesta por los valores aleatorios independientes  $x_1, \dots, x_n$ , la probabilidad, de que cualquier posible trayectoria conste precisamente de estos  $n$  valores discretos, está dada por el producto:

$$\lambda(\theta) = p(x_1, \theta) p(x_2, \theta) \cdots p(x_n, \theta)$$

Mientras, en el caso de variables continuas, la probabilidad de que la observación conste de  $n$  valores en intervalos pequeños  $x_1 \leq x \leq x_1 + \Delta x, \dots, x_n \leq x \leq x_n + \Delta x$  está dada por la expresión:

$$\lambda(\Delta x, \theta)^n = p(x_1, \theta) \Delta x \cdots p(x_n, \theta) \Delta x = \Delta x^n \prod_i^n p(x_i, \theta) \quad (2.67)$$

El valor de  $\lambda$  se denomina *función de verosimilitud* que depende de cada trayectoria  $x_1, \dots, x_n$  y del parámetro desconocido  $\theta$ . En el método de verosimilitud se escoge la aproximación del valor desconocido de  $\theta$ , para un valor de  $\lambda$  tan grande como sea posible. Si  $\lambda$  es una función derivable de  $\theta$ , entonces, una condición necesaria para que  $\lambda$  tenga un máximo está en que

$$d\lambda/d\theta = 0$$

Cualquier solución de la anterior ecuación corresponde a la *estimación de máxima verosimilitud* para el parámetro  $\theta$ , en la cual se reemplazan los valores de la observación por cada una de las variables independientes aleatorias  $\xi_1, \dots, \xi_n$  para obtener la variable aleatoria  $\tilde{\theta}$  denominada *estimador de máxima verosimilitud* con respecto a  $\theta$ . Cuando la FDP incluye  $r$  parámetros desconocidos  $\theta_1, \dots, \theta_r$ , es preferible conformar un sistema de  $r$  ecuaciones en la forma:

$$\frac{d \ln(\lambda)}{d\theta_1} = 0, \dots, \frac{d \ln(\lambda)}{d\theta_r} = 0, \quad (2.68)$$

La ventaja de esta transformación está que las derivadas de los productos se reemplazan por la derivación de sumas. No obstante, en algunos casos, la solución del sistema es difícil de lograr.

Las estimaciones de máxima verosimilitud tienen las siguientes propiedades:

- Cuando existe la estimación suficiente  $\tilde{\theta}(\xi_1, \xi_2, \dots, \xi_N)$  para el parámetro  $\theta$ , entonces, cada solución de la ecuación de verosimilitud es función de  $\tilde{\theta}(\xi_1, \xi_2, \dots, \xi_N)$ .
- Si para  $\theta$  existe una estimación efectiva  $\tilde{\theta}(\xi_1, \xi_2, \dots, \xi_N)$ , entonces, la ecuación de verosimilitud, (2.68) tiene una única solución:  $\tilde{\theta}(\xi_1, \xi_2, \dots, \xi_n)$ .

**Teorema 2.6.** Sea la FDP  $p(\theta, \xi)$ , tal que cumpla las siguientes condiciones:

1. Para casi todos los  $\xi$  existen las derivadas:

$$\frac{\partial^k \ln p(\theta, \xi)}{\partial \theta^k}, \quad k = 1, 2, 3;$$

2. Se cumple la desigualdad:

$$\left| \frac{\partial^k \ln p(\theta, \xi)}{\partial \theta^k} \right| \leq G_k(\xi), \quad k = 1, 2, 3;$$

siendo las funciones  $G_k(\xi)$ ,  $k = 1, 2$  integrables, esto es,

$$\sup_{\theta} \int_{\Xi} G_k(\xi) p(\theta, \xi) d\xi < \infty$$

3. Para cada  $\theta$  la siguiente integral

$$I(\theta) = \int_{\Xi} \left( \frac{\partial \ln p(\theta, \xi)}{\partial \theta} \right)^2 p(\theta, \xi) d\xi \quad (2.69)$$

es positiva y acotada. Entonces, la ecuación de verosimilitud (2.68) tiene solución  $\tilde{\theta}(\xi_1, \dots, \xi_N)$  que es una estimación consistente y asintóticamente efectiva.

**Ejemplo 2.21.** Sea la trayectoria de valores aleatorios estadísticamente independientes,  $\{10, 11, 8, 12, 9\}$ , de un proceso que se asume con FDP exponencial,  $p(\theta, \xi) = \frac{1}{\theta} \exp(-\xi/\theta) u(\xi)$ ,  $\theta > 0$ . Hallar la estimación de máxima verosimilitud para  $\theta$ .

La función de verosimilitud de  $\theta$  es igual a

$$\begin{aligned} \lambda(\theta) &= \prod_{n=1}^N p(\theta, \xi_n) = \prod_{n=1}^5 \frac{1}{\theta} \exp\left(-\frac{\xi_n}{\theta}\right) \\ &= \theta^{-5} \exp\left(-\frac{50}{\theta}\right), \quad \theta > 0 \end{aligned}$$

Al derivar la anterior expresión por el parámetro  $\theta$  se obtiene:

$$\frac{d\lambda}{d\theta} = -5\theta^{-6} \exp\left(-\frac{50}{\theta}\right) + \theta^{-5} \left(\frac{50}{\theta^2}\right) \exp\left(-\frac{50}{\theta}\right)$$

Cuando se hace la derivada igual a 0, al resolver por  $\tilde{\theta}$ , se obtiene el valor concreto de la estimación:  $\tilde{\theta} = 50/5 = 10$ .

En el caso de tener una variable aleatoria  $\xi$  con FDP normal, con valores desconocidos de media y varianza, la función de verosimilitud se determina por la expresión:

$$\lambda = \frac{1}{\sqrt{2\pi}\sigma_{\xi}} \cdots \frac{1}{\sqrt{2\pi}\sigma_{\xi}} e^{-\frac{(\xi_1 - m_{1\xi})^2}{2\sigma_{\xi}^2} - \cdots - \frac{(\xi_N - m_{1\xi})^2}{2\sigma_{\xi}^2}} = \left( \frac{1}{\sqrt{2\pi}\sigma_{\xi}} \right)^{N/2} e^{-h}, \quad (2.70)$$

donde  $h = -\frac{1}{2\sigma_\xi^2} \sum_{k=1}^N (\xi_k - m_{1\xi})^2$ .

Hallando el logaritmo se tiene:

$$A_\xi = \ln(\lambda) = -\frac{N}{2} \ln(\sqrt{2\pi}) - \frac{N}{2} \ln(\sigma_\xi^2) - h \quad (2.71)$$

El máximo del funcional  $\ln(\lambda)$  por el parámetro  $m_{1\xi}$  corresponde a hallar el mínimo del último término (dependiente del parámetro desconocido  $m_{1\xi}$ ), que contiene la suma de los cuadrados  $\sum_{k=1}^N (\xi_k - m_{1\xi})^2$ . Por lo tanto, en el caso concreto de la FDP normal, el método de máxima verosimilitud coincide con el método de los mínimos cuadrados, mostrado en el ejemplo 2.23. Así, para hallar la estimación de la media, se deriva e iguala a cero la ecuación (2.71) por el parámetro desconocido  $m_{1\xi}$ ,

$$\frac{d \ln(\lambda)}{dm_{1\xi}} = \frac{1}{\sigma_\xi^2} \sum_{k=1}^N (\xi_k - m_{1\xi}) = 0,$$

entonces,

$$\sum_{k=1}^N (\xi_k - m_{1\xi}) = \sum_{k=1}^N \xi_k - N m_{1\xi} = 0$$

Al resolver para  $m_{1\xi}$ , expresando la solución en términos de  $\tilde{m}_{1\xi}$ , se obtiene,

$$\tilde{m}_{1\xi} = \frac{1}{N} \sum_{k=1}^N \xi_k \quad (2.72)$$

De igual manera, se obtiene la ecuación para la estimación de la varianza,

$$\frac{d \ln(\lambda)}{d\sigma_\xi^2} = -\frac{1}{2\tilde{\sigma}_\xi^2} \left( N - \frac{1}{\tilde{\sigma}_\xi^2} \sum_{k=1}^N (\xi_k - \tilde{m}_{1\xi})^2 \right) = 0 \quad (2.73)$$

La solución para  $\sigma_\xi^2$ , en términos de  $\tilde{\sigma}_\xi^2$ , resulta en la respectiva estimación:

$$\tilde{\sigma}_\xi^2 = \frac{1}{N} \sum_{k=1}^N (\xi_k - \tilde{m}_{1\xi})^2$$

**Método del mínimo  $\chi^2$ .** Sean  $\xi_1, \xi_2, \dots, \xi_N$  las observaciones independientes correspondientes a la variable aleatoria  $\xi$ , que contienen el parámetro desconocido  $\theta$  y están dadas sobre un espacio  $X$ , tal que pueda ser dividido en  $r$  conjuntos incongruentes de la forma,  $X_1, X_2, \dots, X_r$ . Se asume que la cantidad de observaciones de la trayectoria  $\xi_1, \xi_2, \dots, \xi_N$ , que se localizan dentro del conjunto  $\xi_i$  corresponde a  $N_i$ . Si el conjunto  $X$  es finito (o sea, la variable aleatoria puede tomar una cantidad finita de valores), entonces, se puede considerar que cada  $X_i$  es un conjunto conformado por un solo punto.



De esta manera, se realiza la agrupación de los resultados de la observación, formando la siguiente variable aleatoria (2.31),

$$\chi^2 = \sum_{i=1}^N \frac{(N_i - Np_i(\theta))^2}{Np_i(\theta)}$$

donde  $p_i(\theta) = p_\theta(X_i)$  ( $i = 1, 2, \dots, r$ ). En este caso, la estimación  $\tilde{\theta}(\xi_1, \xi_2, \dots, \xi_N)$  se denomina *estimación por el método del mínimo  $\chi^2$* , la cual se obtiene al minimizar por  $\theta$  la magnitud  $\chi^2$ . Asumiendo que  $\theta$  es un parámetro con dimensión  $d$ , entonces, para encontrar la estimación por el método del mínimo  $\chi^2$  se conforma el sistema de ecuaciones:

$$\sum_{i=1}^N \left( \frac{N_i - Np_i(\theta)}{p_i(\theta)} + \frac{(N_i - Np_i(\theta))^2}{2Np_i(\theta)} \right) \frac{\partial p_i(\theta)}{\partial \theta^k} = 0, \quad k = 1, 2, \dots, d$$

Cabe decir que las propiedades asintóticas de las estimaciones del método del mínimo  $\chi^2$  son muy cercanas a las estimaciones de máxima verosimilitud.

### 2.2.2. Intervalos de confianza

En algunos casos, es importante además de la estimación del parámetro desconocido, determinar la región donde se presume se encuentra el verdadero valor del parámetro. Aunque esa región se construye sobre la observación, sus valores cambian de una a otra trayectoria en forma aleatoria. Por lo tanto, se puede definir la probabilidad con que en una región concreta se localice el verdadero valor del parámetro. Escogiendo un cierto valor suficientemente pequeño  $\alpha > 0$ , que corresponde al valor de significación, se puede construir una regla que permita para cada trayectoria estimar la región de medida, en la cual con una probabilidad de  $1 - \alpha$  el valor del parámetro se encuentra. Esta región se denomina *intervalo de confianza*, mientras el valor  $1 - \alpha$  es el *coeficiente o nivel de confianza*.

Sean  $x_1, x_2, \dots, x_n$  las observaciones independientes de la variable aleatoria  $\xi$ , cuya densidad contiene el parámetro desconocido  $\theta \in \mathbb{R}$ , para el cual se tiene la estimación  $\tilde{\theta}(x_1, x_2, \dots, x_n)$ . Sea  $q_\theta(dt)$  una densidad de la estimación  $\tilde{\theta}$ , suponiendo que el valor verdadero del parámetro coincide con  $\tilde{\theta}$ , esto es,

$$q_\theta(dt) = Q_\theta \left\{ \tilde{\theta}(x_1, x_2, \dots, x_n) \in dt \right\}$$

donde  $Q_\theta(dx_1, dx_2, \dots, dx_n) \in \mathbb{R}^n$  es la distribución, que se determina por:

$$Q_\theta(dx_1, dx_2, \dots, dx_n) = P_\theta(dx_1) P_\theta(dx_2) \cdots P_\theta(dx_n) \quad (2.74)$$

Sea  $q_\theta(dt)$  una densidad sin átomos, entonces para un valor dado de  $\alpha$  se pueden escoger los valores  $a_1(\theta, \alpha)$ ,  $a_2(\theta, \alpha)$ , donde  $a_1(\theta, \alpha) < a_2(\theta, \alpha)$ , para los cuales se cumple,

$$\int_{\{t < a_1(\theta, \alpha)\}} q_\theta(dt) + \int_{\{t > a_2(\theta, \alpha)\}} q_\theta(dt) = \alpha$$

Este método de selección no tiene solución única. De otra parte, asumiendo la continuidad por  $\theta$  de las funciones a escoger y que cada una de las ecuaciones  $a_i(\theta, \alpha) = \lambda$  ( $i = 1, 2$ ) tenga solución única  $c_i(\lambda, \alpha)$  ( $i = 1, 2$ ), entonces, las relaciones son equivalentes:

$$\begin{aligned} Q_\theta \{a_1(\theta, \alpha) < \tilde{\theta} < a_2(\theta, \alpha)\} &= 1 - \alpha \\ Q_\theta \{c_1(\tilde{\theta}, \alpha) < \theta < c_2(\tilde{\theta}, \alpha)\} &= 1 - \alpha \end{aligned}$$

De esta manera, conociendo la distribución de la estimación  $\tilde{\theta}(x_1, x_2, \dots, x_n)$ , para un valor dado de  $\alpha > 0$  se puede construir el intervalo de confianza entre los valores  $(c_1(\tilde{\theta}, \alpha), c_2(\tilde{\theta}, \alpha))$  para un nivel de confianza  $1 - \alpha$ .

Si  $q_\theta(dt)$  tiene átomos, los valores  $a_1(\theta, \alpha)$  y  $a_2(\theta, \alpha)$  se escogen de la desigualdad:

$$\int_{\{t < a_1(\theta, \alpha)\}} q_\theta(dt) + \int_{\{t < a_2(\theta, \alpha)\}} q_\theta(dt) \leq \alpha$$

La desigualdad está dada, por cuanto, no pueden existir valores de  $a_1$  y  $a_2$ , para los cuales tenga lugar la igualdad [17]. Seguidamente, el proceso es similar al descrito anteriormente para la obtención del intervalo de confianza. En forma general, al tomar como base diversas estimaciones  $\tilde{\theta}$  pueden darse diferentes intervalos de confianza, pero en cualquier caso, es importante que su longitud sea lo más pequeña posible. Por esto, en su construcción se emplean estimaciones efectivas o, por lo menos, asintóticamente efectivas que sean obtenidas, por ejemplo del método de máxima verosimilitud.

La construcción del intervalo de confianza asume una distribución  $P_\theta(dx)$  sin átomos y que se puede hallar una función  $g(\theta, x_1, x_2, \dots, x_n)$ ,  $x_i \in \xi$ , con las siguientes propiedades:

- La función  $g(\theta, x_1, x_2, \dots, x_n)$ , es continua y monótona con relación a  $\theta$ .
- La función de distribución  $Q_\theta\{g(\theta, x_1, x_2, \dots, x_n) < \gamma\}$  no depende de  $\theta$ ,  $\forall \gamma \in \mathbb{R}$ .

Para un valor dado de  $\alpha > 0$ , se escogen los valores  $a_1(\alpha)$  y  $a_2(\alpha)$ , tales que:

$$Q_\theta\{a_1(\alpha) < g(\theta, x_1, x_2, \dots, x_n) < a_2(\alpha)\} = 1 - \alpha$$

De acuerdo con la propiedad b, los valores  $a_i(\alpha)$  ( $i = 1, 2$ ) no dependen de  $\theta$ . Al notar por  $c_i$  ( $i = 1, 2$ ) los valores que satisfacen las relaciones  $g(\theta, x_1, x_2, \dots, x_n) = a_i(\alpha)$ , respectivas, que dependen de la observación  $\theta, x_1, x_2, \dots, x_n$  y  $\alpha$ , se obtiene que

$$Q_\theta(c_1 < \theta < c_2) = 1 - \alpha$$

por lo tanto,  $(c_1, c_2)$  corresponde al intervalo de confianza con nivel de confianza  $1 - \alpha$ .

Así por ejemplo, sea dada la distribución integral  $F_\theta(x)$  continua y monótona para  $\theta$ ,

$$F_\theta(x) = \int_{-\infty}^x p_\theta(d\xi), \quad x \in \mathbb{R}$$

Entonces, se puede comprobar que la función  $\prod_{k=1}^n F_\theta(x_k)$  cumple las condiciones  $a$  y  $b$ , por lo que esta se puede emplear para la construcción de los intervalos de confianza. Debido a la continuidad de  $F_\theta(\xi)$ :

$$Q_\theta \{F_\theta(x_k) < \gamma\} = \begin{cases} 0, & \gamma \leq 0 \\ \gamma, & 0 \leq \gamma \leq 1 \\ 1, & \gamma > 1 \end{cases}$$

donde la variable correspondiente a la suma  $\sum_{k=1}^n \log F_\theta(x_k)$  tiene FDP con  $\Gamma$ -densidad:

$$Q_\theta \left\{ -\log a_2 < -\sum_{k=1}^n \log F_\theta(x_k) < -\log a_1 \right\} = \int_{-\log a_2}^{-\log a_1} \frac{1}{\Gamma(n)} \lambda^{n-1} e^{-\lambda} d\lambda$$

Se pueden escoger los valores  $a_1$  y  $a_2$  ( $a_1 < a_2$ ) para un valor dado  $\alpha > 0$ , de tal manera que la integral de la derecha en la última igualdad sea igual a  $1 - \alpha$ , obteniéndose que:

$$Q_\theta \left\{ a_1 < \prod_{k=1}^n F_\theta(x_k) < a_2 \right\} = 1 - \alpha$$

Por cuanto, la función  $\prod_{k=1}^n F_\theta(x_k)$  es monótona por  $\theta$ , entonces, existen ciertos valores  $c_1$  y  $c_2$ , que dependen solamente de  $\alpha$  y  $x_1, \dots, x_n$ , tales que determinen el intervalo de confianza  $(c_1, c_2)$  para  $\theta$  con nivel de confianza  $1 - \alpha$ .

En la práctica, es frecuente el uso de los *percentiles* de nivel  $1 - \alpha$ , entendidos como el valor  $x_{1-\alpha}$  para el cual  $F_\theta(x_{1-\alpha}) = 1 - \alpha$ . Así por ejemplo, en el caso de la FDP normal estándar,  $\mathcal{N}(0, 1)$ , el percentil de nivel  $\alpha = 1 - \varepsilon$  se calcula como:

$$\Phi(\xi_\alpha) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\xi_\alpha} e^{-\lambda^2/2} d\lambda = \alpha$$

**Estimación Bayesiana de intervalos.** Sea el parámetro aleatorio  $\theta$  con FDP a priori  $p(\theta)$ , cuya estimación  $\tilde{\theta}(x_1, \dots, x_n)$  está construida sobre la trayectoria de valores independientes  $x_1, \dots, x_n$ . Se asume que la distribución  $Q_\theta(dt)$  de la estimación  $\tilde{\theta}$  es absolutamente continua, como también lo es la respectiva función  $g(\theta, t)$  que en este caso se toma igual a la FDP condicional definida en (2.41), la cual de acuerdo con (2.4), se determina como,

$$g(\theta, \tilde{\theta}) = p(\theta|\tilde{\theta}) = \frac{p(\theta)p(\tilde{\theta}|\theta)}{\int_{\Theta} p(\theta)p(\tilde{\theta}|\theta)d\theta}$$

donde  $p(\tilde{\theta}|\theta)$  es la FDP condicional de la estimación  $\tilde{\theta}$  para un valor fijo de  $\theta$ , con lo cual la probabilidad de que el parámetro  $\theta$  esté localizado dentro de los límites  $(c_1, c_2)$  está dada

por la expresión,

$$P(c_1 \leq \theta \leq c_2 | \tilde{\theta}) = \int_{c_1}^{c_2} p(\theta | \tilde{\theta}) d\theta$$

Luego, dado el valor de  $\alpha$  se pueden determinar los valores límites,  $c_1(\tilde{\theta}, \alpha)$  y  $c_2(\tilde{\theta}, \alpha)$ , para los cuales se cumpla la igualdad:

$$P(c_1(\tilde{\theta}, \alpha) \leq \theta \leq c_2(\tilde{\theta}, \alpha) | \tilde{\theta}) = 1 - \alpha$$

### 2.2.3. Estimación de parámetros en la distribución normal

**Estimación de la media cuando se conoce la varianza.** Sea el conjunto de valores independientes  $x_1, \dots, x_n$ , de una trayectoria que se considera suficientemente grande, así que  $n > 30$ , correspondientes a la variable aleatoria normal  $x \in \xi$  con FDP, descrita como  $p(\theta, \xi) = \mathcal{N}(\theta, \sigma_\xi^2)$ , donde  $\theta$  es el parámetro desconocido, mientras  $\sigma$  conocida.

Sea la estimación de la esperanza matemática obtenida por el método de los momentos en (2.66),  $k = 1$ . De los resultados obtenidos en el ejemplo 2.20, se deduce que  $\tilde{\theta}$  es una variable aleatoria con distribución normal y momentos:

$$\begin{aligned} \theta(x_1, \dots, x_n) &= \tilde{m}_{1\xi} = m_{1\xi}, \\ \tilde{\sigma}_\xi^2 &= \sigma_\xi^2 / n \end{aligned}$$

luego, la estimación  $\tilde{\theta}$  no presenta sesgo, es consistente y, adicionalmente, efectiva [17]. En concordancia con lo anterior, la FDP es Gaussiana con parámetros  $p(\theta, \xi) = \mathcal{N}(m_{1\xi}, \sigma_\xi^2)$ . De esta manera, para un valor dado  $\alpha > 0$ , teniendo en cuenta (2.74) se deben hallar los valores ( $c_2 = -c_1 = c_\alpha$ ), tales que se cumpla la igualdad:

$$\begin{aligned} Q_\theta \{ \tilde{m}_{1\xi} - c_\alpha < \xi < \tilde{m}_{1\xi} + c_\alpha \} &= P \{ \tilde{m}_{1\xi} - c_\alpha < \xi < \tilde{m}_{1\xi} + c_\alpha \} = P \{ |\xi - \tilde{m}_{1\xi}| < c \} \\ &= \alpha \end{aligned}$$

Teniendo en cuenta la definición de FDP para la variable  $x \in \xi$  se obtiene:

$$\begin{aligned} Q_\theta(dx_1, \dots, dx_n) &= (2\pi\sigma_\xi^2)^{-\frac{n}{2}} \exp \left\{ -\frac{1}{2\sigma_\xi^2} \sum_{k=1}^n (x_k - \theta)^2 \right\} (dx_1, \dots, dx_n) \\ &= \int_{\tilde{m}_{1\xi} - c_\alpha}^{\tilde{m}_{1\xi} + c_\alpha} p_\xi(x) dx = 2\Phi \left( \frac{c_\alpha}{\sqrt{\sigma_\xi^2}} \right) - 1 = 2\Phi \left( c \sqrt{\frac{n}{\tilde{\sigma}_\xi^2}} \right) - 1 \\ &= \alpha \end{aligned}$$

siendo  $\Phi(x)$  la integral de Laplace, definida en (2.27). Así, la distancia obtenida, entre los puntos  $(\tilde{m}_{1\xi} - c, \tilde{m}_{1\xi} + c)$ , corresponde al intervalo de confianza con nivel  $1 - \alpha$ .

**Estimación de la varianza para una media conocida.** En este caso,

$$p(\theta, \xi) = \frac{1}{\sqrt{2\pi\theta}} \exp \left\{ -\frac{(\xi - m_{1\xi})^2}{2\theta} \right\}, \quad \xi \in \mathbb{R}$$

donde la varianza de la estimación  $\tilde{\theta}$ , de acuerdo con el ejemplo 2.20, se determina por la expresión  $E_{\theta}\{\tilde{\theta} - \theta\}^2 = \theta^2/n$ , entonces, se puede demostrar, que el valor  $n\tilde{\sigma}_{\xi}^2/\sigma_{\xi}^2$  tiene  $\chi^2$ -FDP con  $n$  grados de libertad.

El intervalo de confianza con nivel  $1 - \alpha$  es el cálculo de los valores  $a_1$  y  $a_2$ , tales que:

$$Q_{\theta} \left\{ a_1 < n\tilde{\sigma}_{\xi}^2/\sigma_{\xi}^2 < a_2 \right\} = 1 - \alpha$$

donde  $Q_{\theta}(dx_1, \dots, dx_n) = (2\pi\theta)^{-n/2} \exp \left\{ \frac{1}{2\theta} \sum_{k=1}^n (x_k - m_{\xi})^2 \right\} (dx_1, \dots, dx_n)$ . Luego,

$$\begin{aligned} Q_{\theta} \left( n \frac{\tilde{\theta}}{a_2} < \theta < n \frac{\tilde{\theta}}{a_1} \right) &= Q_{\theta} \left( \frac{n\tilde{\sigma}_{\xi}^2}{a_2} < \sigma_{\xi}^2 < \frac{n\tilde{\sigma}_{\xi}^2}{a_1} \right) \\ &= 1 - \alpha \end{aligned}$$

tal que  $(n\tilde{\sigma}_{\xi}^2/a_2, n\tilde{\sigma}_{\xi}^2/a_1)$  es el intervalo de confianza buscado.

**Estimación de la varianza cuando se conoce la media.** Sea la estimación  $\tilde{\theta} = \tilde{m}_{1\xi}$  sin sesgo y consistente. La construcción del intervalo de confianza parte del hecho de que la variable  $(\tilde{m}_{1\xi} - \theta)/\tilde{\mu}_{2\xi}$ , donde  $\tilde{\mu}_{2\xi}$  está dada en el ejemplo 2.20 ( $k = 2$ ), tiene FDP del tipo *Student* con  $n - 1$  grados de libertad, descrita como

$$p_{n-1}(x) = \frac{\Gamma\left(\frac{n}{2}\right)}{\sqrt{n}\Gamma\left(\frac{(n-1)}{2}\right)} (1+x^2)^{-\frac{n}{2}}$$

Con lo cual, el valor  $c_{\alpha}$  se determina de la relación

$$2 \int_0^{c_{\alpha}} p_{n-1}(x) dx = 1 - \alpha$$

Por lo que se obtiene  $Q_{\theta} \left\{ -c_{\alpha} < \frac{\tilde{m}_{1\xi} - \theta}{\sqrt{\tilde{\mu}_{2\xi}}} < c_{\alpha} \right\} = 1 - \alpha$ , donde

$$Q_{\theta}(dx_1, \dots, dx_n) = (2\pi\sigma_{\xi}^2)^{-n/2} \exp \left\{ -\frac{1}{2\sigma_{\xi}^2} \sum_{k=1}^n (x_k - \theta)^2 \right\} (dx_1, \dots, dx_n)$$

De esta manera, el intervalo de confianza para la estimación del parámetro  $\theta$  con nivel  $1 - \alpha$  corresponde al segmento  $(\tilde{m}_{1\xi} - \sqrt{\tilde{\mu}_{2\xi}}c_{\alpha}, \tilde{m}_{1\xi} + \sqrt{\tilde{\mu}_{2\xi}}c_{\alpha})$ .

**Estimación conjunta de la varianza y la media.** Sean desconocidos los parámetros de la FDP normal,  $m_{1\xi}$  y  $\sigma_\xi^2$ , estimados, bien sea por el método de los momentos o el de máxima verosimilitud:

$$\tilde{m}_{1\xi} = \frac{1}{n} \sum_{k=1}^n x_k, \quad \tilde{\sigma}_\xi^2 = \tilde{\mu}_{2\xi} = \frac{1}{n} \sum_{k=1}^n (x_k - \tilde{m}_{1\xi})^2$$

En este caso, la estimación de  $\sigma_\xi^2$  resulta sesgada, por lo que hay necesidad de introducir el factor  $n/n - 1$ , para que la corrección del sesgo  $\tilde{\sigma}_\xi^2 = n\tilde{\sigma}_\xi^2/n - 1$ , entonces,  $(\tilde{m}_{1\xi}, \tilde{\sigma}_\xi^2)$  corresponden a las respectivas estimaciones no sesgadas de los parámetros  $(m_{1\xi}, \sigma_\xi^2)$ .

#### 2.2.4. Prueba de hipótesis

Sea una estructura estadística relacionada con la variable  $x \in \xi$ , para la cual la función de densidad de probabilidad  $p_\xi(x)$  es desconocida. Cualquier suposición que defina unívocamente a una clase de pertenencia la densidad  $p_\xi(x)$  se denomina *hipótesis* [18]. Una hipótesis  $H$  se determina como *simple* si el conjunto  $H$  consiste de un único parámetro  $\theta$ . En caso contrario,  $H$  es una hipótesis *compuesta*. En la práctica, se hace necesario la verificación en la correspondencia entre los resultados reales del experimento con la hipótesis que se tiene sobre su estructura aleatoria. En este sentido, se realiza el procedimiento de verificación o *prueba de hipótesis*, el cual permite de las observaciones analizadas tomar o rechazar una hipótesis dada.

La prueba estadística de hipótesis pertenece a la clase de estimación puntual y se realiza dividiendo el espacio de observaciones  $x \in \xi$  en dos espacios que no se intercepten:  $\mathfrak{X}_0 \cap \mathfrak{X}_1 = \emptyset$ , correspondientes a las dos hipótesis alternativas. Si los resultados de las observaciones pertenecen a  $x \in \mathfrak{X}_0$  (intervalo de toma de hipótesis), entonces se considera que la respectiva hipótesis se confirma por los datos empíricos y, por lo tanto, se acepta  $H_0$  como cierta. En caso contrario, cuando  $x \notin \mathfrak{X}_0$  (intervalo crítico), entonces, la hipótesis dada  $H_0$  no corresponde a las observaciones experimentales y, por lo tanto, se rechaza. Por cuanto, los valores de cada trayectoria de  $x \in \xi$  son aleatorios, se definen las respectivas probabilidades a priori de acierto de la prueba de la hipótesis  $P_\theta(x \in \mathfrak{X}_0)$  y  $P_\theta(x \notin \mathfrak{X}_0)$ , dado un valor de  $\theta$ . Sin embargo, la misma prueba de hipótesis puede generar errores de primer género (rechazar una hipótesis cuando realmente era cierta) y de segundo género (tomar la hipótesis como cierta cuando realmente era falsa), entonces, el procedimiento de prueba se debe optimizar para obtener el valor mínimo de ambos géneros de error.

Cabe anotar que, en la mayoría de los casos prácticos, es imposible optimizar el procedimiento de la verificación de la hipótesis, en el sentido de minimizar ambos géneros de error para cualquier valor tan pequeño como se quiera [17].

El valor máximo permitido  $\alpha$  del error de primer género para un criterio dado se denomina *valor de significación del criterio*,  $\sup_{\theta \in H} P_\theta(X_0) \leq \alpha$ . En el caso del error de segundo género, en vez de analizar su valor, notado por  $\beta$ , se analiza la magnitud,  $w = 1 - \beta$ , denominada *potencia del criterio* de la prueba de hipótesis  $H_0$  contra la hipótesis alternativa  $H_1$

La comparación de los diferentes criterios propuestos se realiza empleando métricas de efectividad asintóticas, basadas en el estudio de la convergencia de la función de potencia en las cercanías del parámetro  $\theta \in H$ . Se considera óptimo aquel criterio, que para un valor dado de significación  $\alpha$ , brinde el máximo valor de potencia del criterio, el cual define la siguiente función crítica  $\varphi(x)$ :

$$\int_{\xi} \varphi(x) p_0(x) dx = \alpha, \quad \varphi(x) = \begin{cases} 1, & p_1(x) \geq \lambda_{\alpha} p_0(x) \\ 0, & p_1(x) < \lambda_{\alpha} p_0(x) \end{cases} \quad (2.75)$$

siendo  $p_0(x)$  y  $p_1(x)$  las respectivas FDP de  $P_0$  y  $P_1$ .

En la práctica, para el desarrollo del criterio (2.75), denominado de *Neyman-Pearson*, en calidad de estadística suficiente se emplea la relación de verosimilitud (2.67), que implica la verificación de la desigualdad  $p_1(x) \geq \lambda_{\alpha} p_0(x)$  para una observación concreta  $x \in \xi$ . Si la desigualdad se cumple, entonces la hipótesis  $H_0$  se rechaza, en caso contrario se acepta. La constante  $\lambda_{\alpha} \in \mathbb{R}$  se determina a partir de la condición:  $P(\Lambda_x \geq \lambda_{\alpha}) = \alpha$ , siendo  $\Lambda_x = p_1(x)/p_0(x)$  la respectiva relación de verosimilitud.

**Ejemplo 2.22.** Dada la trayectoria de la variable aleatoria Gaussiana  $x \in \xi$  de longitud  $n$  y varianza conocida  $\sigma_{\xi}^2$ , realizar la prueba de hipótesis  $H_0$  sobre el valor medio de una variable aleatoria,  $m_{1\xi} = m_0$  contra la hipótesis alternativa  $H_1$  sobre el valor  $m_{1\xi} = m_1$  para el criterio de Neyman-Pearson empleando la estadística suficiente (2.67).

La función de verosimilitud del parámetro de distribución  $m_{1\xi}$  para ambas hipótesis, teniendo en cuenta (2.70), se determina como

$$\lambda(m_i; x) = (2\pi\sigma_{\xi}^2)^{-n/2} \exp\left(-\frac{1}{2\sigma_{\xi}^2} \sum_{k=1}^n (x_k - m_i)^2\right), \quad i = 0, 1$$

con lo cual la relación de verosimilitud es igual a

$$\Lambda_x = \frac{m_1 - m_0}{\sigma_{\xi}^2} \sum_{k=1}^n x_k - \frac{n(m_1^2 - m_0^2)}{2\sigma_{\xi}^2},$$

cuyo valor se compara con el umbral  $h_0 = \ln h$ , en concordancia con (2.71). La hipótesis nula se toma en favor de  $m_{1\xi} = m_0$  y se acepta como cierta si se cumple que,

$$\frac{m_1 - m_0}{\sigma_{\xi}^2} \sum_{k=1}^n x_k - \frac{n(m_1^2 - m_0^2)}{2\sigma_{\xi}^2} \underset{H_0}{\overset{H_1}{\gtrless}} \ln h \quad (1)$$

La desigualdad (1) se puede escribir en formas de las siguientes desigualdades, que dependen de la relación entre los valores  $m_0$  y  $m_1$ :

$$\begin{aligned} m_0 < m_1 : \quad \tilde{m}_{1\xi} &= \frac{1}{n} \sum_{k=1}^n x_k < \frac{m_1 - m_0}{2} + \frac{\sigma_{\xi}^2 \ln h}{n(m_1 - m_0)} = h_1 \\ m_0 > m_1 : \quad \tilde{m}_{1\xi} &= \frac{1}{n} \sum_{k=1}^n x_k > \frac{m_1 - m_0}{2} + \frac{\sigma_{\xi}^2 \ln h}{n(m_1 - m_0)} = h_2 \end{aligned}$$

De las anteriores desigualdades, se observa que para tomar la decisión sobre el valor medio es necesario comparar su valor estimado con los umbrales  $h_i, i = 1, 2$ , los cuales se escogen de acuerdo al criterio de optimización empleado.

### 2.2.5. Estimación de dependencias funcionales

Sea  $\xi \in \mathbb{R}^p$  y  $\eta \in \mathbb{R}^q$  dos vectores aleatorios dependientes. Se exige establecer la dependencia entre  $\xi$  y  $\eta$ , a partir de los resultados de observación  $(x_1, x_2, \dots, x_n)$  para el vector  $\xi$  y de la observación  $(y_1, y_2, \dots, y_n)$  del vector  $\eta$ .

Sea  $m(y_k, y_l)$  alguna métrica en  $\mathbb{R}^q$ . Se debe hallar una función

$$f(\cdot) : \mathbb{R}^p \rightarrow \mathbb{R}^q$$

para la cual  $E\{m(\eta, f(\xi))\}$  toma su mínimo valor. En principio, se puede escoger la norma Euclídea, descrita en (1.2), tal que

$$m(y_k, y_l) = \|y_k - y_l\|_2 \text{ en } \mathbb{R}^q$$

por lo que la solución se da en términos de la *función teórica de regresión*  $y = f(x)$ , descrita como

$$f(x) = E\{\eta | \xi = x\} \quad (2.76)$$

donde  $x$  se denomina la *variable de regresión*, mientras  $y$  es la *respuesta*.

La solución de (2.76) exige el conocimiento, por lo menos a nivel de estimación, sobre la función de densidad de probabilidad conjunta de los vectores  $\xi$  y  $\eta$ , lo cual en la práctica no es real. En este sentido, se pueden tomar las siguientes aproximaciones de solución:

- En calidad de estimación de la función teórica de regresión,  $y = f(x)$ , se define una función de aproximación de la clase  $\tilde{f} = \{f : \mathbb{R}^p \rightarrow \mathbb{R}^q\}$
- Se asume que la función  $f(x) = f(x, \theta)$ , de manera unívoca se determina por una cierta cantidad de parámetros  $\theta \in \mathbb{R}^m$ , siendo  $m$  un valor fijo. Por cierto, se asume además la dependencia lineal entre  $f(x, \theta)$  y  $\theta$ .

En consideración con lo anterior, resulta el siguiente modelo lineal de regresión:

$$y = \theta_0 f_0(x) + \theta_1 f_1(x) + \dots + \theta_{m-1} f_{m-1}(x) \quad (2.77)$$

donde  $f_i(x)$  son funciones conocidas que se determinan por la naturaleza del experimento,  $m$  es el *orden del modelo de regresión* y  $\theta = (\theta_0, \theta_1, \dots, \theta_{m-1})$  corresponde al vector de parámetros a ser estimado sobre la base de los resultados de las observaciones  $(x_1, x_2, \dots, x_n)$  y  $(y_1, y_2, \dots, y_n)$ , teniendo como restricción la desigualdad  $m < n$ .

Los siguientes modelos son los que mayor aceptación tienen en el análisis de regresión estadística:



1. *Modelos de aproximación empíricos* [17, 19].

$$\begin{aligned}
 y &= \theta_0 + \theta_1 x, & y &= \sum_k \theta_k x^k, & y &= \theta_0 + \theta_1 \ln x \\
 y &= \theta_0 + \theta_1 x + \theta_2 \frac{1}{x}, & \ln y &= \sum_k \theta_k x^k, & \ln y &= \theta_0 + \theta_1 \ln x, \\
 1/y &= \sum_k \theta_k x^k, & y &= \theta_0 + \theta_1 / \sqrt{x} & y &= \theta_0 + \theta_1 10^{\theta_2} \ln x
 \end{aligned}$$

2. *Funciones de descomposición lineal*. Usualmente, se emplean las funciones ortogonales en la representación del tipo (1.4), particularmente, las funciones de Chebyshev.

3. *Funciones de ajuste (spline)*. Cuando el modelo de regresión se da por intervalos mediante polinomios de interpolación o alguna función con estructura a priori dada.

El modelo lineal es, tal vez, el de más amplio empleo en los métodos de regresión estadística:

$$\mathbf{y} = \boldsymbol{\theta} \mathbf{x} + \boldsymbol{\varepsilon} \quad (2.78)$$

siendo  $\boldsymbol{\varepsilon}$  el vector aleatorio de errores de observación. En general, se supone que no hay errores sistemáticos de medida y, por lo tanto, el valor medio de la variable aleatoria es  $E\{\boldsymbol{\varepsilon}_k\} = 0$ , para  $k = 1, \dots, n$ , con potencia desconocida de error de observación, definida como  $E\{\boldsymbol{\varepsilon}_k, \boldsymbol{\varepsilon}_l\} = \sigma_\varepsilon^2 \delta_{kl}$ .

En caso de asumir que los errores de medición les corresponde la correlación expresada en la forma  $E\{\boldsymbol{\varepsilon}, \boldsymbol{\varepsilon}^\top\} = \sigma_\varepsilon^2 \boldsymbol{\Sigma}$ , siendo conocida la matriz  $\boldsymbol{\Sigma}$ , se puede realizar el cambio de variables:  $\boldsymbol{\lambda}' = \boldsymbol{\lambda}^{-1/2} \boldsymbol{\lambda}$ , donde  $\boldsymbol{\lambda} = \{\mathbf{y}, \mathbf{x}, \boldsymbol{\varepsilon}\}$ , que conlleva al modelo con momentos en la forma  $E\{\boldsymbol{\varepsilon}'\} = 0$  y  $E\{\boldsymbol{\varepsilon}', \boldsymbol{\varepsilon}'^\top\} = \sigma_\varepsilon^2 \mathbf{I}$ , donde  $\mathbf{I}$  es la matriz unidad, en este caso con dimensión  $n \times n$ .

Si en el modelo (2.78)  $q = 1$  y el orden de regresión  $m$  es conocido, la estimación  $\tilde{\boldsymbol{\theta}}$  del vector de parámetros  $\boldsymbol{\theta}$  del modelo puede realizarse por diferentes métodos. Sin embargo, el más empleado corresponde al *método de mínimos cuadrados*:

$$\min_{\boldsymbol{\theta}} \sum_{k=1}^n \left[ y_i - \sum_{l=0}^{m-1} f_l(x_k) \theta_l \right]^2 = \sum_{k=1}^n \left[ y_i - \sum_{l=0}^{m-1} f_l(x_k) \tilde{\theta}_l \right]^2 \quad (2.79)$$

Sea el hiperespacio lineal en  $\mathbb{R}^m$  conformado por los valores del parámetro  $\boldsymbol{\theta}$ , cuya estimación  $\tilde{\boldsymbol{\theta}}$  es obtenida por (2.79), entonces se tendrá la siguiente interpretación geométrica:

- si  $\boldsymbol{\theta} \in \mathbb{R}^m$  y  $\text{rank}\{\mathbf{x}\} = m$ , entonces el vector  $\mathbf{x}\tilde{\boldsymbol{\theta}}$  corresponde a la proyección del vector de observaciones  $\mathbf{y} \in \mathbb{R}^n$ ,  $m < n$  en el hiperespacio  $\boldsymbol{\theta} \in \mathbb{R}^m : \mathbf{x}\tilde{\boldsymbol{\theta}} = \text{pr}_{\boldsymbol{\theta}} \mathbf{y}$ . La estimación  $\tilde{\boldsymbol{\theta}}$  será única.
- si  $\boldsymbol{\theta} \in \mathbb{R}^m$ ,  $\text{rank}\{\mathbf{x}\} < m$  y el vector  $\text{pr}_{\boldsymbol{\theta}} \mathbf{y}$  no pertenece al espacio de valores de la transformación lineal  $\mathcal{G}\{\mathbf{x}\}$ , entonces el vector  $\mathbf{x}\tilde{\boldsymbol{\theta}}$  corresponde a la proyección del

vector  $\text{pr}_\theta \mathbf{y}$ . La estimación  $\tilde{\theta}$  será única.

- si  $\theta \in \mathbb{R}^m$ ,  $\text{rank}\{\mathbf{x}\} < m$  y  $\text{pr}_\theta \mathbf{y} \in \mathcal{G}\{\mathbf{x}\}$ , entonces el mínimo en (2.79) se obtiene en todo vector  $\tilde{\theta} = \tilde{\theta}' + \mathbf{h}$ , donde  $\tilde{\theta}'$  es cualquier vector del hiperespacio  $\theta \in \mathbb{R}^m$ , ortogonal a  $\mathcal{G}\{\mathbf{x}\}$ . La estimación  $\tilde{\theta}$  por mínimos cuadrados no es única.

El empleo del método de mínimos cuadrados para la obtención de las estimaciones puntuales de los parámetros desconocidos de los modelos lineales de regresión, a diferencia de otros métodos de estimación, dados en el numeral §2.2.1, no exige información a priori sobre el tipo de FDP, la cual muchas veces no se tiene en las etapas iniciales de proceso de datos.

---

**Ejemplo 2.23.** Sea la relación  $\mathbf{y} = f(\mathbf{x})$ , que deber ser confirmada a partir de un conjunto dado de  $n$  observaciones, tales que cada elemento  $y_i$  de la función  $\mathbf{y}$  o *respuesta* es obtenido para un valor fijo  $x_i$ , correspondiente a la variable  $\mathbf{x}$  o *factor*. En el caso de aceptar la hipótesis de dependencia lineal o *análisis de regresión lineal*, se deben hallar los parámetros de la recta  $\mathbf{y} = a + b\mathbf{x}$ , que para un criterio de error de representación dado, generalmente el error cuadrático medio (2.79), se alcance el valor mínimo (principio de Lagrange):

$$\min \sum_{i=1}^n (y_i (a - bx_i))^2, \mathbf{x} \in (\mathbf{x}_{\min}, \mathbf{x}_{\max}), \mathbf{y} \in (\mathbf{y}_{\min}, \mathbf{y}_{\max})$$

De tal manera, que la recta obtenida permite, con cierto valor de probabilidad, pronosticar el valor de  $\mathbf{y}$  para un valor de  $\mathbf{x}$ , dadas las observaciones  $\{x_i, y_i\}$ .

Sean los resultados de medida  $y_i$  distribuidos normalmente,  $p(\mathbf{y}) \sim \mathcal{N}(\sigma_{\mathbf{y}}, y_i)$ , tal que la probabilidad de obtener el valor  $y_i$  sea igual a:

$$P_{a,b}(y_i) \sim \frac{1}{\sigma_{\mathbf{y}}} \exp(-y_i - a - bx_i)^2 / 2\sigma_{\mathbf{y}}^2$$

la cual depende de los coeficientes  $a$  y  $b$ . Por lo tanto, la probabilidad de obtener el conjunto de resultados de medida  $y_1, \dots, y_n$ , asumiendo su independencia estadística, es igual a

$$P_{a,b}(y_1, \dots, y_n) = P_{a,b}(y_1) \dots P_{a,b}(y_n) \sim \frac{1}{\sigma_{\mathbf{y}}^n} \exp(-\chi^2/2)$$

siendo  $\chi^2 = \sum_{i=1}^n (y_i - a - bx_i)^2 / \sigma_{\mathbf{y}}^2$ .

La estimación de las constantes  $a$  y  $b$ , por el método de máximo de probabilidad a posteriori corresponde a la mayor probabilidad de  $P_{a,b}(y_1, \dots, y_n)$ , esto es, cuando la suma  $\chi^2$  es mínima o método de *mínimos cuadrados*, descrito por la expresión (2.79). El valor óptimo se halla diferenciando parcialmente la suma por las variables  $a$  y  $b$ , e igualando a cero:

$$\begin{aligned} \frac{\partial \chi^2}{\partial a} &= \left(-2/\sigma_{\mathbf{y}}^2\right) \sum_{i=1}^n (y_i - a - bx_i) = 0 \\ \frac{\partial \chi^2}{\partial b} &= (-2\sigma_{\mathbf{y}}^2) \sum_{i=1}^n x_i (y_i - a - bx_i) = 0 \end{aligned}$$

Las anteriores expresiones se pueden representar como un sistema de ecuaciones, teniendo como variables  $a$  y  $b$  (ecuaciones normales), de la forma:

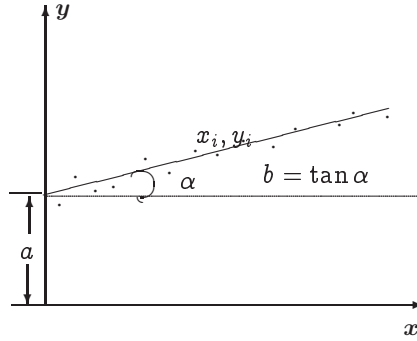
$$\begin{aligned} an + b \sum_{i=1}^n x_i &= \sum_{i=1}^n y_i \\ a \sum_{i=1}^n x_i + b \sum_{i=1}^n x_i^2 &= \sum_{i=1}^n x_i y_i \end{aligned}$$

por lo que se obtiene ( $\Sigma$  denota la suma por  $i = 1, \dots, n$ ):

$$a = \frac{(\Sigma x_i^2)(\Sigma y_i) - (\Sigma x_i)(\Sigma x_i y_i)}{\Delta} \quad (1a)$$

$$b = \frac{n(\Sigma x_i y_i) - (\Sigma x_i)(\Sigma y_i)}{\Delta} \quad (1b)$$

siendo  $\Delta = n(\Sigma x_i^2) - (\Sigma x_i)^2$ .



**Figura 2.10.** Interpretación geométrica de la regresión lineal

Geoméricamente, el coeficiente  $a$ , obtenido en (1a), corresponde a la distancia desde el origen de las coordenadas hasta el punto, donde se intercepta la línea de regresión con la ordenada, mientras el coeficiente  $b$ , obtenido en (1b), corresponde a la tangente del ángulo  $\alpha$  de inclinación de la línea de regresión, como se muestra en la Figura 2.10.

**Ejemplo 2.24.** Realizar la prueba de hipótesis sobre la validez de ecuación de regresión lineal simple obtenida en el ejemplo 2.23.

Se asume que el valor que pondera qué tanto el conjunto de resultados de medida, contenidos en las parejas  $\{(x_i, y_i) : i = 1, \dots, n\}$ , confirma la hipótesis sobre la dependencia lineal entre  $\mathbf{x}$  e  $\mathbf{y}$ , corresponde al coeficiente de correlación dado en (2.46),

$$\rho = \frac{\sigma_{xy}}{\sigma_x \sigma_y} \quad (2)$$

donde la suma,  $\sigma_{xy} = \frac{1}{n} \sum_{i=1}^n (x_i - \tilde{m}_{1x})(y_i - \tilde{m}_{1y})$  es un término paralelo al segundo momento centralizado de una variable dado en (2.9), definida como la *desviación combinada*, pero para la cual no se cumplen todas las condiciones (2.14). Reemplazando la anterior definición en (2) y teniendo en cuenta la estimación de la varianza dada en (2.73), se obtiene la

estimación del coeficiente de correlación lineal en la forma

$$\tilde{\rho} = \frac{\sum_i (x_i - \tilde{m}_{1x})(y_i - \tilde{m}_{1y})}{\left( \sum_i (x_i - \tilde{m}_{1x})^2 \sum_i (y_i - \tilde{m}_{1y})^2 \right)^{1/2}} \quad (3)$$

Por cuanto,  $-1 \leq \rho \leq 1$ , si  $|\tilde{\rho}| \rightarrow 1$ , entonces los puntos están ubicados en las cercanías de una línea recta. En cambio, si  $|\tilde{\rho}| \rightarrow 0$ , los puntos no están correlacionados y, por lo tanto, no se agrupan sobre ninguna recta. A partir de coeficiente de correlación (3), el modelo de la regresión ortogonal se construye por la expresión:

$$\mathbf{y} = \tilde{m}_{1y} + \frac{2\tilde{\rho}}{\sigma_0 + \sqrt{\sigma_0^2 + 4\tilde{\rho}^2}} (\mathbf{x} - \tilde{m}_{1x})$$

la cual está directamente relacionada con el concepto de la elipse de dispersión, mostrada en la Figura 2.8(b).

La estimación de los coeficientes de regresión es analizada estadísticamente, mediante la prueba de hipótesis  $H_0 : \rho = 0$ , esto es, si hay diferencia o no, desde el punto de vista estadístico, en la estimación del coeficiente de regresión con respecto al valor 0 [19]. Se puede demostrar que cuando  $\rho = 0$ , la estadística  $\tilde{\rho}$  se describe mediante la  $\Gamma$ -densidad (2.28), de la forma [20]:

$$\frac{\Gamma\{\frac{1}{2}(n-1)\}}{\Gamma\{\frac{1}{2}(n-2)\}\sqrt{\pi}} (1 - \tilde{\rho}^2)^{(n-4)/2}$$

que es equivalente a la distribución de Student (2.30) con  $n - 2$  grados de libertad para la variable aleatoria

$$(n-2)^{1/2} \frac{\tilde{\rho}}{\sqrt{(1 - \tilde{\rho}^2)}}$$

El caso más importante de análisis es la prueba de hipótesis  $H_0 : \rho = 0$  contra  $H_1 : \rho \neq 0$ . La hipótesis  $H_0$  se rechaza con un nivel de significación  $\alpha$ . Por lo anterior, los límites del intervalo de análisis se fijan:

$$\hat{t} = \left| \tilde{b} \right| / \tilde{\sigma}_b \geq t_{1-\frac{1}{2}\alpha} \{n-2\} \quad (4)$$

siendo  $\tilde{b}$  la estimación del coeficiente de regresión. Si la desigualdad (4) se cumple, entonces se puede decir que  $\tilde{\rho}$  significativamente se diferencia de cero.

La estimación del coeficiente de correlación (prueba de existencia sobre la correlación), se realiza de manera similar:

$$\hat{t} = \frac{\tilde{\rho}\sqrt{n-2}}{1 - \tilde{\rho}^2} \geq t_{1-\frac{1}{2}\alpha} \{n-2\}$$

Si la anterior condición se cumple, entonces la hipótesis  $H_0 : \tilde{\rho} = 0$ , se rechaza.

La prueba completa sobre la validez de ecuación de regresión, se efectúa mediante la  $F$ -estadística

de Fisher (2.32):

$$\tilde{F} = \frac{\frac{1}{k-2} \sum_{i=1}^k n_i (\tilde{m}_{1y_i} - \tilde{y}_i)^2}{\frac{1}{n-k} \sum_{i=1}^k \sum_{l=1}^{n_i} (y_{il} - \tilde{m}_{1y})^2} \quad (5)$$

donde  $\tilde{y}_i = a + b\tilde{x}_i$ . El criterio implica que si la suma de los valores medios de las desviaciones de grupo de la regresión directa, dividida sobre la suma de las desviaciones de los valores  $\mathbf{y}$  con respecto a los valores medios de grupo, alcanza o excede los límites de significación, entonces la hipótesis sobre linealidad hay que rechazarla. La estimación (5) exige que la longitud del vector  $\mathbf{y}$  sea mayor, en  $k$  valores a la respectiva longitud del vector  $\mathbf{x}$ , esto es, que a cada valor  $x_i$  le corresponde  $n_i$  valores del vector  $\mathbf{y}$ , lo que en la práctica, frecuentemente ocurre.

Si la hipótesis sobre la linealidad no puede ser tomada, entonces se pueden realizar pruebas sobre modelos no lineales, en primera instancia, de tipo polinomial:

$$y = a + bx + cx^2 + \dots$$

Aunque su implementación exige el empleo de métodos computacionales. Sin embargo, después de cierto orden, la varianza empieza a crecer desmesuradamente, por lo que el algoritmo no siempre converge.

En general, empleando el método de los mínimos cuadrados (2.79), se puede construir, prácticamente, cualquier forma de dependencia no lineal. En este caso, se emplea alguna transformación de linealización, por cuanto, solamente las funciones lineales pueden ser reconstruidas por el método de mínimos cuadrados.

---

## Problemas

**Problema 2.12.** Dada una observación de valores independientes con volumen  $n$ , constatar que la estimación del parámetro  $\lambda$  de la FDP exponencial, calculada por cualquiera de los métodos de los momentos o de máximo verosimilitud, converge en el mismo valor. Calcular la varianza de la estimación y analizar su efectividad.

**Problema 2.13.** Dado el valor del coeficiente de confianza,  $1 - \alpha = .95$ , hallar los valores extremos del intervalo de confianza del parámetro  $\lambda$  de una variable aleatoria con FDP exponencial, dada una observación independiente de valores con volumen  $n$ .

**Problema 2.14.** Dada una observación de valores independientes con volumen  $n$ , mediante el método de máxima verosimilitud hallar la estimación del valor medio de una variable aleatoria con FDP Rayleigh, descrita por (1) en el ejemplo 2.15.

**Problema 2.15.** Probar la hipótesis nula  $H_0$  sobre el valor de la varianza,  $\sigma_\xi^2 = k_0$ , de una variable aleatoria Gaussiana, con volumen  $n$  y valor  $m_{1\xi}$ , contra la hipótesis alternativa  $H_1$  sobre el valor  $\sigma_\xi^2 = k_1$ , siendo  $k_0 < k_1$ .

**Problema 2.16.** Basados en el criterio de Neyman-Pearson realizar la prueba de hipótesis sobre el tipo de FDP con parámetros conocidos, para un valor dado de significación  $\alpha$ . Sea la hipótesis nula,  $p(x|H_0) = (2\pi)^{-1/2} \exp(-x^2/2)$ , la FDP normal, mientras la hipótesis alternativa corresponde a la densidad exponencial simétrica  $p(x|H_1) = (1/2) \exp(-|x|)$ .

**Problema 2.17.** Un dosímetro, durante un tiempo  $T = 10s$ , registra la cantidad  $k = 20$  partículas. Realizar la prueba de hipótesis mediante el criterio de Neyman-Pearson para un valor dado de significación  $\alpha$ , siendo la hipótesis nula  $H_0$ , que considera que el resultado de la observación está condicionado por la radiación de fondo, contra la hipótesis alternativa  $H_1$  que corresponde a la suma del fondo mas una fuente de radiación. Los valores medios de partículas por segundo son conocidos e iguales a  $n_0 = 0.2$  y  $n_1 = 0.8$ , de forma correspondiente.

**Problema 2.18.** Probar la hipótesis nula  $H_0$  sobre el valor del parámetro  $\lambda = k_0$  de una variable aleatoria con FDP exponencial con volumen  $n$ , contra la hipótesis alternativa  $H_1$  sobre el valor  $\lambda = k_1$ , siendo  $k_0 < k_1$ .

**Problema 2.19.** Dadas dos observaciones independientes de variables aleatorias Gaussianas, con volúmenes  $n_1 = 16$  y  $n_2 = 25$ , se obtiene las correspondientes estimaciones del coeficiente de correlación  $r_1 = 0.5$  y  $r_2 = 0.7$ . Realizar la prueba de hipótesis, con valor de significación  $\alpha = 0.05$ , por la cual ambas observaciones corresponden a un mismo conjunto general.

**Problema 2.20.** Estimar los coeficientes de la regresión lineal  $y = \tilde{a} + \tilde{b}x$ , dadas las observaciones,

$x_i$	6.4	8.0	7.4	6.8	7.2	6.7	7.7	8.2
$y_i$	2.2	5.2	4.8	3.3	4.0	3.0	4.2	5.3

Realizar la prueba de hipótesis sobre la validez de ecuación de regresión lineal simple obtenida y estimar el intervalo de confianza para un valor dado  $1 - \alpha = 0.9$ .

## 2.3. Teoría de decisión estadística

La estimación óptima de procesos aleatorios está basada en la teoría de decisión estadística, en la medida en que la información extraída del proceso aleatorio se emplea para la obtención de algún objetivo de análisis, cuyo resultado es una decisión.

Aunque la variedad de tareas, en las cuales se hace necesario la toma de decisión, es amplia se tiene una formulación matemática con las siguientes propiedades:

- (a). Cualquier solución está orientada a la obtención de un objetivo concreto, que bien puede estar definido por una cantidad, finita o infinita, de variantes con sus respectivas bondades y restricciones, sobre las cuales se estima la calidad de la decisión a tomar.
- (b). La decisión se toma, basados tanto en la información como en la evidencia estadística a priori y obtenidos durante el proceso de solución.
- (c). La toma de decisión dentro de un conjunto de posibles variantes puede contemplar la incertidumbre, determinada por la respectiva probabilidad de toma de cada decisión.

La decisión seleccionada dentro de todas las posibles variantes, que brinde la mayor bondad de solución de acuerdo a un objetivo propuesto, se denomina *óptima*.

### 2.3.1. Definiciones básicas

**Funciones de decisión.** Sea la tarea de estimación de procesos aleatorios, en la cual se tiene la trayectoria de un vector aleatorio  $\mathbf{y}(t) \in \mathfrak{Y}$ . Así mismo, sea  $\mathbf{x}(t) \in \mathfrak{X}$  un proceso aleatorio vectorial, cuyas trayectorias no son conocidas al observador y que corresponden al valor verdadero del proceso, pero cuya información está contenida en los resultados de observación  $\mathbf{y}(t)$ .

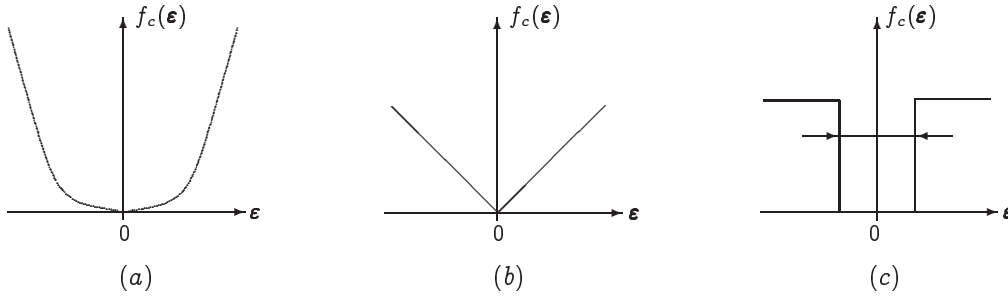
Sea el vector  $\mathbf{D} \in \mathfrak{D}$  de cada posible solución, que se puede tomar con respecto a  $\mathbf{x}(t)$  por los resultados de observación  $\mathbf{y}(t)$ , esto es, cada solución establece una relación entre los puntos de  $\mathfrak{Y}$ . Sea  $\mathbf{g} \in \mathcal{G}$  la transformación de la forma  $\mathbf{D} = \mathbf{g}(\mathbf{y})$ .

Las funciones de decisión pueden ser, tanto determinísticas cuando la transformación no contiene ninguna medida de incertidumbre, como randomizada cuando la transformación implica que para cada  $\mathbf{y}$ , cualquier  $\mathbf{D}$  se toma con una medida de probabilidad condicional  $p(\mathbf{D}|\mathbf{y})$ .

**Funciones de pérdida.** Como resultado de la toma de una u otra decisión pueden ocurrir errores, por lo cual se establece una medida cuantitativa de la respectiva ganancia o pérdida, en forma de dependencia  $f_c\{\mathbf{x}, \mathbf{d}\}$  denominada *función de pérdida*, que se selecciona en concordancia con la naturaleza física de la tarea a resolver. Debido a que la solución  $\mathbf{D} = \mathbf{g}(\mathbf{y})$  depende de la trayectoria del proceso aleatorio  $\mathbf{y}(t)$ , entonces el valor actual de pérdida  $f_c\{\mathbf{x}, \mathbf{g}(\mathbf{y})\}$  es también aleatorio. Por lo tanto, las funciones de decisión es debido seleccionarlas de acuerdo a la comparación de las características de aleatoriedad de la función de pérdida. Así por ejemplo, para el caso de la tarea de estimación, la decisión  $\mathbf{D}$

representa la estimación  $\tilde{\mathbf{x}}(t)$  que corresponde al vector proceso aleatorio  $\mathbf{x}(t)$ . Al introducir la medida de error  $\boldsymbol{\varepsilon}(t) = \mathbf{x}(t) - \tilde{\mathbf{x}}(t)$ , entonces, la función de pérdida se describe como  $f_c\{\mathbf{x}, \tilde{\mathbf{x}}\} \triangleq f_c\{\boldsymbol{\varepsilon}\}$ . Por cierto, a la función de pérdidas se denomina *permitida* si presenta las siguientes propiedades:

- (a).  $f_c\{\boldsymbol{\varepsilon}\}$  es función escalar de  $n$  variables, correspondientes a la dimensión del proceso  $\mathbf{x}(t)$ ,
- (b).  $f_c\{\boldsymbol{\varepsilon} = 0\} = 0$ ,
- (c).  $f_c\{\boldsymbol{\varepsilon}\} = f_c\{-\boldsymbol{\varepsilon}\}$ , la función de pérdida es simétrica, en el sentido en que no importa en que dirección se tenga el error, la pérdida se considera la misma.
- (d).  $f_c\{\boldsymbol{\varepsilon}_m\} > f_c\{\boldsymbol{\varepsilon}_n\}$ ,  $\forall \rho(\boldsymbol{\varepsilon}_m) \geq \rho(\boldsymbol{\varepsilon}_n)$ , donde  $\boldsymbol{\varepsilon}_m$  y  $\boldsymbol{\varepsilon}_n$  son los valores de error de estimación,  $\rho$  es una función escalar positiva semidefinida convexa, con argumento de  $n$  variables, y corresponde a la medida de distancia del error  $\boldsymbol{\varepsilon}$ , desde el origen de las coordenadas de un espacio euclídeo con dimensión  $n$ . La función  $f_c\{\boldsymbol{\varepsilon}\}$  es monótona ascendente con respecto al incremento del argumento  $\boldsymbol{\varepsilon}$ .



**Figura 2.11.** Funciones de pérdida.

En la práctica, amplio uso tienen las siguientes tres formas de funciones permitidas de pérdida:

a. Cuadrática, Figura 2.11(a),  $f_c(\boldsymbol{\varepsilon}) = \|\boldsymbol{\varepsilon}\|_{\mathbf{B}}^2$  (2.80a)

b. Por módulo, Figura 2.11(b),  $f_c(\boldsymbol{\varepsilon}) = \|\boldsymbol{\varepsilon}\|_{\mathbf{B}}$  (2.80b)

c. Simple, Figura 2.11(c),  $f_c(\boldsymbol{\varepsilon}) = \begin{cases} 0, & \|\boldsymbol{\varepsilon}\|_{\mathbf{B}} < e/2 \\ 1/e, & \|\boldsymbol{\varepsilon}\|_{\mathbf{B}} \geq e/2 \end{cases}$  (2.80c)

donde  $\|\cdot\|$  es la norma del vector,  $e$  es una constante y  $\mathbf{B}$  es una matriz positiva semidefinida de dimensión  $n \times n$ , compuesta por los elementos  $b_{ij}$  que son factores de peso. En la estimación de señales es usual el empleo de la norma del error

$$\|\boldsymbol{\varepsilon}_{\mathbf{B}}\|^2 = \boldsymbol{\varepsilon}^T \mathbf{B} \boldsymbol{\varepsilon}$$



**Funciones de riesgo.** Como anteriormente se dijo, los valores concretos de la función de pérdida  $f_c(\mathbf{x}, \mathbf{D})$  son aleatorios, y por lo tanto, las funciones de decisión se seleccionan basados en la comparación de las propiedades estadísticas de la función de pérdida, la cual se denomina *función de riesgo* (o riesgo, simplemente) y corresponde a las pérdidas que en promedio se obtienen sobre el conjunto de decisiones posibles.

La selección de la decisión óptima se lleva a cabo mediante la minimización de los riesgos por todas las posibles funciones de decisión. En la teoría de toma de decisiones, en calidad de criterio de calidad generalizado, se emplea el riesgo medio, que para el caso de las funciones de decisión *randomizadas* corresponde a la esperanza matemática de la función de pérdidas en la forma,

$$R = E \{f_c\{\mathbf{x}, \mathbf{D}\}\} = \iiint f_c\{\mathbf{x}, \mathbf{D}\} p(\mathbf{D}|\mathbf{y}) p(\mathbf{x}, \mathbf{y}) d\mathbf{D} d\mathbf{y} d\mathbf{x} \quad (2.81)$$

donde  $p(\mathbf{D}|\mathbf{y})$  es la densidad de probabilidad condicional de la toma de decisión  $\mathbf{D}$  para un valor dado del vector  $\mathbf{y}$ , que caracteriza la función de decisión,  $p(\mathbf{x}, \mathbf{y})$  es la FDP conjunta de los vectores  $\mathbf{x}$  e  $\mathbf{y}$ . Teniendo en cuenta que

$$p(\mathbf{x}, \mathbf{y}) = p(\mathbf{x})p(\mathbf{y}|\mathbf{x}) = p(\mathbf{y})p(\mathbf{x}|\mathbf{y})$$

entonces, el riesgo medio (2.81) se puede representar en forma alterna como

$$R = \iiint f_c(\mathbf{x}, \mathbf{D}) p(\mathbf{D}|\mathbf{y}) p(\mathbf{x}, \mathbf{y}) p(\mathbf{y}) d\mathbf{D} d\mathbf{y} d\mathbf{x} \quad (2.82a)$$

$$R = \iiint f_c(\mathbf{x}, \mathbf{D}) p(\mathbf{D}|\mathbf{y}) p(\mathbf{y}, \mathbf{x}) p(\mathbf{x}) d\mathbf{D} d\mathbf{y} d\mathbf{x} \quad (2.82b)$$

donde  $p(\mathbf{x}|\mathbf{y})$  es la FDP condicional del proceso no observado  $\mathbf{x}(t)$  para un valor dado  $\mathbf{y}(t)$ , además,  $p(\mathbf{x})$  y  $p(\mathbf{y})$ , son la FDP marginales de  $\mathbf{x}$  e  $\mathbf{y}$ , respectivamente, mientras,  $p(\mathbf{y}|\mathbf{x})$  es la FDP condicional del proceso observado  $\mathbf{y}$  para una valor de  $\mathbf{x}$ , que es la función de verosimilitud de  $\mathbf{x}$ .

Las expresiones (2.81), (2.82a), y (2.82b) caracterizan las pérdidas en promedio por todas las posibles decisiones  $\mathbf{D} \in \mathfrak{D}$ , así como todos los posibles valores de  $\mathbf{x}$  e  $\mathbf{y}$ . La optimización de la función de decisión consiste en la selección de aquella función  $p(\mathbf{D}|\mathbf{y})$  que brinde el mínimo riesgo medio, determinado por la función (2.81).

Cuando se analizan funciones *no randomizadas* de decisión, a cada  $\mathbf{y}$  le corresponde una determinada solución  $\mathbf{D} = \mathbf{g}(\mathbf{y})$  y la FDP condicional es  $p(\mathbf{D}|\mathbf{y}) = \delta(\mathbf{D} - \mathbf{g}(\mathbf{y}))$ , con lo cual el riesgo medio es igual a,

$$R_g = E \{f_c\{\mathbf{x}, \mathbf{g}(\mathbf{y})\}\} = \iint f_c\{\mathbf{x}, \mathbf{g}(\mathbf{y})\} p(\mathbf{x}, \mathbf{y}) d\mathbf{y} d\mathbf{x} \quad (2.83)$$

En general, el riesgo medio (2.83) es una función no lineal con relación a la función de decisión  $\mathbf{g}(\mathbf{y})$ . Las respectivas funciones (2.82a) y (2.82b) del riesgo medio para el caso de

las funciones de decisión no randomizadas tienen la forma,

$$R_g = \iint f_c(\mathbf{x}, \mathbf{g}(\mathbf{y})) p(\mathbf{x}|\mathbf{y}) p(\mathbf{y}) d\mathbf{y} d\mathbf{x} \quad (2.84a)$$

$$R_g = \iint f_c(\mathbf{x}, \mathbf{g}(\mathbf{y})) p(\mathbf{y}|\mathbf{x}) p(\mathbf{x}) d\mathbf{y} d\mathbf{x} \quad (2.84b)$$

de los cuales se puede determinar el riesgo condicional  $r_g(\mathbf{x})$  por  $\mathbf{x}$ , que corresponde al promedio de la función de pérdida por la función de verosimilitud  $p(\mathbf{y}|\mathbf{x})$  para un valor dado de  $\mathbf{x}$ ,

$$r_g(\mathbf{x}) = \int f_c(\mathbf{x}, \mathbf{g}(\mathbf{y})) p(\mathbf{y}|\mathbf{x}) d\mathbf{y} \quad (2.85)$$

El promedio del riesgo condicional  $r_g(\mathbf{x})$  por todos los posibles valores de  $\mathbf{x}$ , de acuerdo a (2.84b), resulta en el riesgo medio,

$$R_g = \int r_g(\mathbf{x}) p(\mathbf{x}) d\mathbf{x} \quad (2.86)$$

El riesgo condicional  $r_g(\mathbf{x})$  caracteriza la calidad del sistema que emplea la función de decisión  $\mathbf{g}(\mathbf{y})$ , para un valor dado de  $\mathbf{x}$ . Así que se prefiere aquella función de decisión  $\mathbf{g}(\mathbf{y})$  que presente el menor valor de  $r_g(\mathbf{x})$ . Se define la función de decisión  $\mathbf{g}_1(\mathbf{y})$  como *uniformemente mayor* que  $\mathbf{g}_2(\mathbf{y})$ , cuando se cumple que  $r_{g_1}(\mathbf{x}) \leq r_{g_2}(\mathbf{x}) \forall \mathbf{x} \in \mathfrak{X}$ , además,  $r_{g_1}(\mathbf{x}) < r_{g_2}(\mathbf{x})$  por lo menos para un valor de  $\mathbf{x}$  [21]. De otra parte, la función decisión  $\mathbf{g}$  se denomina *permitida*, si en  $\mathfrak{G}$  no existe otra función decisión  $\mathbf{g}_1$ , que sea uniformemente mayor que  $\mathbf{g}$ .

La clase  $\mathfrak{G}$  de funciones de decisión se denomina *completa*, si para cualquier  $\mathbf{g}_i \notin \mathfrak{G}$  se halla una función de decisión  $\mathbf{g} \in \mathfrak{G}$  uniformemente mayor que  $\mathbf{g}_i$ . En este sentido, si se tiene una clase completa  $\mathfrak{G}$  de funciones de decisión completa, entonces la función de decisión preferible es suficiente con esta clase sin necesidad de analizar las funciones  $\mathbf{g} \notin \mathfrak{G}$  y, por lo tanto, la construcción de clases completas es una tarea importante en la teoría de decisión estadística.

Cuando se tiene alguna función de decisión  $\mathbf{g}(\mathbf{y})$ , que minimice el riesgo condicional  $r_G(\mathbf{x})$ , es suficientemente diferente para los valores de  $\mathbf{x}$ . Por esta razón, la decisión óptima se selecciona basados en diferentes aproximaciones, entre ellas, las de Bayes, minimax, máximo de verosimilitud, etc.

Cuando las observaciones son un conjunto de valores aleatorios, o bien son procesos continuos sobre un intervalo de tiempo de análisis, pero con dimensión alta o infinita, entonces en la teoría de decisión estadística se emplea el concepto de *estadísticas suficientes*, que corresponde a la clase de transformaciones  $\gamma(\mathbf{y})$ , que contienen toda la información, considerada necesaria para la toma de decisión y que está inmersa en las observaciones  $\mathbf{y} \in \mathfrak{Y}$ . Para estas transformaciones la función de decisión óptima  $\mathbf{g}(\gamma(\mathbf{y}))$ , que depende de  $\gamma$ , brinda el mismo valor de riesgo que la decisión óptima  $\mathbf{g}(\mathbf{y})$ . En este caso, la transformación  $\gamma(\mathbf{y})$ , que es la suficiente estadística, presenta importancia para hallar las funciones de forma

de decisión óptimas del tipo, tanto bayesianas, como no bayesianas.

En el caso de la aproximación bayesiana, para la toma de decisión, las estadísticas suficientes corresponden a funciones vectoriales conocidas  $\boldsymbol{\gamma} = \{\gamma_i(\mathbf{y}) : i = 1, \dots, n\}$ , por medio de las cuales la FDP conjunta se puede representar en la forma

$$p(\mathbf{x}, \mathbf{y}) = f_c(\mathbf{y})f(\gamma_1(\mathbf{y}), \dots, \gamma_n(\mathbf{y}), \mathbf{x})$$

donde el factor de multiplicación  $f_c(\mathbf{x})$  no depende de  $\mathbf{x}$ , y la función  $f$  está relacionada con  $\mathbf{y}$ , solamente mediante las funciones  $\gamma_1(\mathbf{y}), \dots, \gamma_n(\mathbf{y})$ .

Como se observa, en este caso  $f$  tiene una cantidad finita de variables. Además, la FDP condicional a posteriori va a depender solamente de  $\gamma_i(\mathbf{y}), \dots, \gamma_n(\mathbf{y})$ ,

$$p(\mathbf{x}|\mathbf{y}) = \frac{f(\gamma_1(\mathbf{y}), \dots, \gamma_n(\mathbf{y}), \mathbf{x})}{\int f(\gamma_1(\mathbf{y}), \dots, \gamma_n(\mathbf{y}), \mathbf{x})d\mathbf{x}}$$

Por lo tanto, las funciones  $\gamma_i(\mathbf{y})$  son suficientes para el cálculo de la FDP a posteriori condicional,

$$p(\mathbf{x}|\mathbf{y}) = p(\mathbf{x}|\gamma_1(\mathbf{y}), \dots, \gamma_n(\mathbf{y}))$$

que al reemplazar en (2.84a) implica que el algoritmo óptimo de estimación también está relacionado con  $\mathbf{y}$  solamente a través de la estadística suficiente,  $\mathbf{g}(\mathbf{y}) = \mathbf{g}(\gamma_1(\mathbf{y}), \dots, \gamma_n(\mathbf{y}))$ .

En la aproximación no bayesiana para la toma de decisión mediante las estadísticas suficientes, de manera similar, la función de verosimilitud tiene la forma,  $p(\mathbf{y}|\mathbf{x}) = f_c(\mathbf{y})$ .

Una estadística suficiente se considera que es más preponderante, entre mayor valor de compresión brinde de los datos observados  $\mathbf{y}(t)$ , esto es, entre menor sea su dimensión de representación. La transformación suficiente de mínima dimensión se denomina *estadística mínima suficiente* para una clase dada de tareas. Como ejemplos de estadísticas mínimas suficientes se tiene la relación de verosimilitud en la tarea de detección binaria de señales, y el funcional de verosimilitud en la estimación de parámetros de las señales.

La tarea más importante de estimación corresponde al riesgo condicional por  $\mathbf{y}$  o riesgo a posteriori  $r_g(\mathbf{y})$ , que es la esperanza media condicional para una función de decisión dada  $\mathbf{g}$  desde la trayectoria de  $\mathbf{y}$ , mientras, el riesgo  $r_g(\mathbf{y})$  se forma por el promedio de la función de pérdida sobre toda la función de densidad de probabilidad a posteriori, para una trayectoria conocida  $\mathbf{y}$ ,

$$r_g(\mathbf{y}) = \int f_c(\mathbf{x}, \mathbf{g}(\mathbf{y}))p(\mathbf{x}|\mathbf{y})d\mathbf{x} \quad (2.87)$$

El riesgo (2.87) determina el valor medio de pérdidas como resultado de la toma de la decisión  $\mathbf{D} = \mathbf{g}(\mathbf{y}) \in \mathfrak{D}$ , que corresponde a una observación para  $\mathbf{y}$ . El riesgo medio  $R_g$ , en concordancia con (2.84a) es el promedio del riesgo a posteriori  $r_g(\mathbf{y})$  por todos los posibles valores de  $\mathbf{y}$ ,

$$R_g = \int r_g(\mathbf{y})p(\mathbf{y})d\mathbf{y} \quad (2.88)$$

### 2.3.2. Decisión Bayesiana

En general, la búsqueda de la función de decisión óptima considera dos casos extremos de análisis sobre el estado de conocimiento a priori para un proceso vectorial no observado  $\mathbf{x}$ : conocimiento completo a priori, cuando se tiene la FDP,  $p(\mathbf{x})$ , y conocimiento incompleto a priori, cuando  $p(\mathbf{x})$  es totalmente desconocida. En el primer caso, se construyen las *funciones de decisión bayesiana*, en el segundo las denominadas *no bayesianas*.

En la aproximación bayesiana se cumplen las dos siguientes condiciones:

1. El proceso  $\mathbf{x}$  es aleatorio y la respectiva FDP marginal  $p(\mathbf{x})$  existe.
2. La FDP a priori  $p(\mathbf{x})$  es conocida al observador

En este caso, la decisión óptima se halla al minimizar el riesgo medio (2.83),

$$\hat{R}_g = \min_g R_g = \min_g \iint f_c(\mathbf{x}, \mathbf{g}(\mathbf{y})) p(\mathbf{x}, \mathbf{y}) d\mathbf{x} d\mathbf{y} \quad (2.89)$$

La función  $\hat{\mathbf{g}}$ , que cumple (2.89), se denomina *función de decisión bayesiana*, al igual que las correspondientes decisión y riesgo medio obtenido. De esta manera, el criterio bayesiano de optimización consiste en la minimización del riesgo medio para una FDP  $p(\mathbf{x})$ , dada a priori. Debido a que el mismo riesgo medio varía en dependencia de la estructura de función de pérdida  $f_c(\mathbf{x}, \mathbf{g}(\mathbf{y}))$ , entonces el criterio de optimización, en forma general, depende de la forma de la función de pérdida. Sin embargo, es conocido que el conjunto total de las decisiones bayesianas, con relación a todas las posibles FDP a priori  $p(\mathbf{x})$ , conforma una clase completa.

Como se observa de (2.88), el procedimiento de minimización del riesgo medio  $R_g$  por  $\mathbf{g}$  implica la minimización por  $\mathbf{g}$  del riesgo a posteriori  $r_g(\mathbf{y})$ , definido en (2.87), que está condicionado por el criterio positivo semidefinido de la FDP  $p(\mathbf{y})$ , lo que implica que el criterio de optimización bayesiano es equivalente al criterio del mínimo del riesgo a posteriori,

$$\hat{r}_g(\mathbf{y}) = \min_g r_g(\mathbf{y}) = \min_g \int f_c(\mathbf{x}, \mathbf{g}(\mathbf{y})) p(\mathbf{x}|\mathbf{y}) d\mathbf{x} \quad (2.90)$$

desde el punto de vista analítico, la tarea (2.90) en algunos casos tiene solución.

En la estimación de procesos aleatorios  $\mathbf{x}$ , la función de pérdida usualmente se determina por un error  $\boldsymbol{\epsilon}$ , cuyo valor pueda ser minimizado seleccionando la estimación  $\tilde{\mathbf{x}}$ . La estimación que cumple esta exigencia se denomina *óptima bayesiana*, o simplemente *óptima* cuando de forma tácita se asume que es bayesiana, la cual en forma general, depende de la selección de la función de pérdida  $f_c(\boldsymbol{\epsilon})$ . Sin embargo, para una clase amplia de tareas, las estimaciones, que son óptimas para una clase de función de pérdida, conservan esta propiedad para la clase completa de funciones de pérdida. Esto es, se puede hablar de invariabilidad de la estimación óptima. En particular, según teorema de Sherman [12] se tiene que si  $f_c(\boldsymbol{\epsilon})$  es una función permitida de pérdida y la FDP a posteriori del proceso  $\mathbf{x}$  es del tipo unimodal y simétrica con relación a la moda, entonces la estimación óptima

bayesiana corresponde a la esperanza matemática condicional del vector  $\mathbf{x}$ . Este caso tiene lugar, por ejemplo, en la estimación lineal de procesos Gaussianos.

---

**Ejemplo 2.25.** Sea la función cuadrática de pérdida (2.80a) y el riesgo bayesiano (2.89). Hallar el valor de la estimación  $\tilde{\mathbf{x}}$  que brinde el menor error para una observación dada  $\mathbf{y}$ .

El riesgo medio bayesiano para el criterio de error dado, tiene la forma

$$R_g = \iint (\mathbf{x} - \tilde{\mathbf{x}})^\top \mathbf{B}(\mathbf{x} - \tilde{\mathbf{x}}) p(\mathbf{x}, \mathbf{y}) d\mathbf{x} d\mathbf{y}$$

Por lo tanto, el riesgo a posteriori tiene la forma

$$r_g(\mathbf{y}) = \int (\mathbf{x} - \tilde{\mathbf{x}})^\top \mathbf{B}(\mathbf{x} - \tilde{\mathbf{x}}) p(\mathbf{x}|\mathbf{y}) d\mathbf{x} \quad (1)$$

donde  $\mathbf{B}$  es una matriz positiva definida, cuyos elementos tienen las unidades de la función de pena por error sobre las unidades del producto de los errores de las respectivas componentes del proceso vectorial estimado  $\mathbf{x}$ . Como se observa en (1), el riesgo a posteriori está relacionado con el momento central de segundo orden de la FDP a posteriori, lo que significa que el mínimo valor de posible riesgo a posteriori, en el caso de la función cuadrática de pérdida, se suple con la estimación  $\tilde{\mathbf{x}}$  con el mínimo valor de la traza de la matriz de momentos centrales escalares de segundo orden (si el proceso es escalar, entonces con el mínimo valor cuadrático medio). En este caso, el valor de la estimación que brinde el menor error cuadrático medio, dada la observación  $\mathbf{y}$ , se obtiene haciendo igual a cero el gradiente  $r_g(\mathbf{y})$  por la variable  $\tilde{\mathbf{x}}$ ,

$$\frac{\partial r_g(\mathbf{y})}{\partial \tilde{\mathbf{x}}} = 2 \int \mathbf{B}(\tilde{\mathbf{y}} - \mathbf{x}) p(\mathbf{x}|\mathbf{y}) d\mathbf{x} = 0 \quad (2)$$

La igualdad (2) se cumple cuando  $\tilde{\mathbf{x}} \int p(\mathbf{x}|\mathbf{y}) d\mathbf{x} = \int \mathbf{x} p(\mathbf{x}|\mathbf{y}) d\mathbf{x}$ .

Asumiendo que se cumple la condición de normalización para  $p(\mathbf{x}|\mathbf{y})$ , entonces se tiene

$$\tilde{\mathbf{x}} = \int \mathbf{x} p(\mathbf{x}|\mathbf{y}) d\mathbf{x} \quad (3)$$

Por lo tanto, la estimación con mínimo error cuadrático medio corresponde al valor de la FDP a posteriori  $p(\mathbf{x}|\mathbf{y})$ . De (3), se observa que la estimación óptima no depende de la matriz  $\mathbf{B}$ , lo que significa se tiene un óptimo general por todas las componentes del proceso  $\mathbf{x}$ .

**Ejemplo 2.26.** En el ejemplo anterior, suponer que se tiene la función simple de pérdida, definida en (2.80c).

En este caso, la minimización del riesgo a posteriori (2.90), que toma la forma

$$r_g(\mathbf{y}) = \frac{1}{\varepsilon} \int p(\mathbf{x}|\mathbf{y}) d\mathbf{x}$$

corresponde a la selección del valor  $\tilde{\mathbf{x}}$ , que maximiza la expresión

$$\frac{1}{\varepsilon} - r_g(\mathbf{y}) = \frac{1}{\varepsilon} - \frac{1}{\varepsilon} \int p(\mathbf{x}|\mathbf{y}) d\mathbf{x} = \frac{1}{\varepsilon} \int p(\mathbf{x}|\mathbf{y}) d\mathbf{x}$$

donde la integración se realiza sobre el espacio  $\mathfrak{E}_1$ , en el cual  $\|\epsilon\|_{\mathbf{B}} \geq \epsilon/2$ , o sobre el espacio  $\mathfrak{E}_0$ , donde se cumple  $\|\epsilon\|_{\mathbf{B}} < \epsilon/2$ . El empleo de la función simple de pérdidas implica un valor relativamente pequeño para el intervalo  $\epsilon$  ( $\epsilon \rightarrow 0$ ). Entonces, para obtener el máximo de la última expresión como valor de la estimación óptima  $\tilde{\mathbf{x}}$ , en cada trayectoria  $\mathbf{y}$  dada, se debe seleccionar el valor de  $\mathbf{x}$ , para el cual  $p(\mathbf{x}|\mathbf{y})$  toma el máximo valor posible, esto es, la estimación  $\tilde{\mathbf{x}}$  corresponde a la moda  $p(\mathbf{x}|\mathbf{y})$ . Esta estimación se denomina *estimación por máximo de densidad de probabilidad a posteriori*.

Conocida la trayectoria  $\mathbf{y}$ , la estimación óptima  $\tilde{\mathbf{x}}$  se halla como la raíz de la ecuación

$$\left. \frac{\partial p(\mathbf{x}|\mathbf{y})}{\partial \mathbf{x}} \right|_{\mathbf{x}=\tilde{\mathbf{x}}} = 0 \quad (4)$$

El valor de la estimación óptima  $\tilde{\mathbf{x}}$  se puede obtener mediante la maximización de la función  $p(\mathbf{x}|\mathbf{y})$  con relación a  $\mathbf{x}$ . Basados en el teorema de Bayes, (2.4), se tiene la relación condicional  $p(\mathbf{x}|\mathbf{y}) = p(\mathbf{x}|\mathbf{y})p(\mathbf{x})/p(\mathbf{y})$ , esto es,  $\ln p(\mathbf{x}|\mathbf{y}) = \ln p(\mathbf{x}, \mathbf{y}) - \ln p(\mathbf{y}) = \ln p(\mathbf{x}|\mathbf{y}) + \ln p(\mathbf{x}) - \ln p(\mathbf{y})$ .

Debido que  $p(\mathbf{y})$  no depende la variable  $\mathbf{x}$ , la maximización de  $p(\mathbf{x}|\mathbf{y})$  se reduce a hallar el valor máximo de la densidad de probabilidad conjunta  $p(\mathbf{x}, \mathbf{y})$ . De esta manera, el valor óptimo de la estimación se puede hallar como la raíz de una de las siguientes ecuaciones:

$$\left. \frac{\partial p(\mathbf{x}, \mathbf{y})}{\partial \mathbf{x}} \right|_{\mathbf{x}=\tilde{\mathbf{x}}} = \left. \frac{\partial p(\mathbf{y}|\mathbf{x})p(\mathbf{x})}{\partial \mathbf{x}} \right|_{\mathbf{x}=\tilde{\mathbf{x}}} = 0 \quad (5a)$$

$$\left. \frac{\partial \ln p(\mathbf{x}, \mathbf{y})}{\partial \mathbf{x}} \right|_{\mathbf{x}=\tilde{\mathbf{x}}} = \left( \left. \frac{\partial \ln p(\mathbf{y}|\mathbf{x})}{\partial \mathbf{x}} + \frac{\partial \ln p(\mathbf{x})}{\partial \mathbf{x}} \right) \right|_{\mathbf{x}=\tilde{\mathbf{x}}} = 0 \quad (5b)$$

La solución de la función de pérdida, en tareas de aplicación concretas, en parte, se realiza atendiendo al principio de implementación simple de los algoritmos óptimos de estimación. Desde este punto de vista, como se observa de (3) y (4), la estimación óptima, para el caso de la función simple de pérdida, se determina por las propiedades locales de FDP a posteriori de las componentes de estimación del vector  $\mathbf{x}$ , en la cercanía de su máximo global, en cambio, cuando se tiene la función cuadrática de pérdida, la estimación se determina por la forma de cambio de la FDP a priori, pero en todo el intervalo del vector  $\mathbf{x}$ . Por esta razón, es de esperar que para algunas tareas de estimación sea preferible la realización del algoritmo (4), que exigen menor detalle de la descripción de la FDP a posteriori, que la requerida para (3), aunque es posible que su rendimiento sea peor.

### 2.3.3. Decisión no Bayesiana

La aproximación Bayesiana implica que sobre el proceso  $\mathbf{x}$ , siendo aleatorio, se conoce su medida probabilística, en particular, su FDP. En caso contrario, se emplean diferentes métodos no Bayesianos de selección de la mejor función de decisión.

**Método de máxima verosimilitud.** Se considera que las mejores estimaciones bayesianas se obtienen por el método del máximo de probabilidad a posteriori, cuando la estimación se calcula mediante la maximización de la FDP conjunta  $p(\mathbf{x}, \mathbf{y}) = p(\mathbf{y}|\mathbf{x})p(\mathbf{x})$ . Sin embargo, en la práctica es usual que no se tenga información a priori alguna sobre el proceso  $\mathbf{x}$ , caso en el cual, es imposible obtener la solución de las ecuaciones (5a) y (5b) del ejemplo 2.26. Por lo tanto, una forma aproximada de la estimación óptima  $\tilde{\mathbf{x}}$  se halla simplificando el sistema de ecuaciones que contengan solo la función de verosimilitud. En este caso, la

estimación obtenida se denomina de *máxima verosimilitud* y se determina por la solución de una de las siguientes ecuaciones,

$$\left. \frac{\partial p(\mathbf{y}|\mathbf{x})}{\partial \mathbf{x}} \right|_{\mathbf{x}=\tilde{\mathbf{x}}} = 0 \quad (2.91a)$$

$$\left. \frac{\partial \ln p(\mathbf{y}|\mathbf{x})}{\partial \mathbf{x}} \right|_{\mathbf{x}=\tilde{\mathbf{x}}} = 0 \quad (2.91b)$$

El método de máxima verosimilitud se puede analizar como un caso particular de la estimación Bayesiana. Así por ejemplo, si en alguna de las expresiones (2.91a) ó (2.91b), se asume que el proceso aleatorio  $\mathbf{x}$  tiene FDP Gaussiana y para un aumento ilimitado de las varianzas de todas las componentes del vector  $\mathbf{x}$ , cuando se cumple que  $\partial \ln p(\mathbf{x})/\partial \mathbf{x} = 0$ , entonces la expresión (5b) se transforma en (2.91b). A propósito, al mismo resultado se llega asumiendo que la FDP es uniforme.

En general, el cambio de (5a) a (2.91a), o bien de (5b) a (2.91b) es aceptable y está condicionado al hecho de que la observación,  $\mathbf{y}$ , contenga tanta información sobre el verdadero valor de  $\mathbf{x}$ , que su estimación  $\tilde{\mathbf{x}}$  tenga de forma significativa una menor dispersión con respecto al valor verdadero  $\mathbf{x}$ , la cual debe ser menor que la dispersión determinada de forma a priori por la densidad  $p(\mathbf{x})$ . Además, se debe tener seguridad que  $p(\mathbf{x})$  no contiene picos muy altos en las regiones de definición de  $\mathbf{x}$ . En algunos casos, cuando la información de la FDP a priori es insuficiente, se asumen aproximaciones Gaussianas o uniformes y se obtienen estimaciones denominadas *seudobayesianas* [12].

El error cuadrático medio de estimación, para el método de máxima verosimilitud, es mayor que para el caso de estimación pseudobayesiana, y por su puesto mucho mayor que para el caso Bayesiano. La razón de este comportamiento del error de estimación está en que la construcción del estimador Bayesiano incluye la información a priori del vector estimado,  $\mathbf{x}$ , mientras en el caso de máxima verosimilitud, ésta se ignora o bien se toma igual para el peor caso.

**Método minimax.** De la expresión (2.86), se observa que a cada vector  $\mathbf{x}$  le corresponde un valor de riesgo condicional  $r_g(\mathbf{x})$ , lo que significa que este valor cambia para cada una de las señales, y por lo tanto, existirá un valor máximo  $r_g(\mathbf{x})_{\max}$  para todos los posibles valores de  $\mathbf{x}$ .

La función de decisión  $\hat{g}(\mathbf{y})$  determina la solución minimax, si se cumple que

$$\sup \hat{r}_g(\mathbf{x}) \leq \sup r_g(\mathbf{x})$$

Asumiendo la existencia de los valores extremos, la anterior relación toma la forma,

$$\min_g \max_{\mathbf{x}} r_g(\mathbf{x}) = \max_{\mathbf{x}} \hat{r}_g(\mathbf{x}) \quad (2.92)$$

El método de estimación minimax, descrito por (2.92), implica que se realiza la minimización por todo  $g$  del riesgo condicional  $r_g(\mathbf{x})$ , para el peor de los casos, por todos los

valores de  $\mathbf{x}$ . Además, la función de decisión  $\hat{g}(\mathbf{x})$  minimiza, por todos los  $\mathbf{x}$ , el máximo valor de la función de riesgo condicional  $r_g(\mathbf{x})$ . La principal particularidad del método (2.92) está en que garantiza un cierto valor de éxito, referido al peor de los casos a ser considerado. Sin embargo, la estimación puede resultar demasiado lejana de las situaciones con mayor probabilidad de ocurrencia.

La solución minimax en muchas tareas de aplicación real significa un procedimiento bastante complejo de resolver. Aunque si se asocia el método minimax al Bayesiano, la complejidad de solución disminuye. Así por ejemplo, la función de decisión minimax  $\hat{g}(\mathbf{x})$ , dadas ciertas restricciones de carácter muy suave (que son simples de cumplir), corresponde a la función Bayesiana con respecto a la distribución a priori menos favorable, que a propósito, frecuentemente resulta ser la FDP uniforme.

**Ejemplo 2.27.** Sea  $x(t)$  un proceso aleatorio escalar, cuyos valores deben localizarse dentro del intervalo de trabajo  $-a < x < a$ ,  $x \in \mathbb{R}$ , en caso contrario, el sistema se considera desintonizado y debe generarse una señal de alarma, por lo menos, mientras el proceso no regrese al intervalo indicado de trabajo. Hallar la estimación óptima  $\tilde{x}(t)$ , dada la trayectoria de observación  $y(t)$  con valores pertenecientes al intervalo  $(t_0, t)$ , que corresponde a una función conocida del proceso estimado mas una perturbación.

La respectiva decisión  $d = g(y)$  es el resultado de la filtración, esto es,  $g(y) = \tilde{x}(t)$ , que se determina por alguna de las funciones de pérdida, por ejemplo, la forma simple (2.80c),

$$f_c(e) = \begin{cases} 0, & |e| \leq \mu \\ c, & |e| > \mu \end{cases}$$

donde  $e(t) = x(t) - \tilde{x}(t)$ ,  $c = \text{const.}$  y  $\mu = \text{const.}$

El sistema puede generar error en dos situaciones: una, cuando la estimación está dentro del intervalo permitido, mientras el valor verdadero no lo está (error de primer tipo), o cuando la estimación está fuera del intervalo, pero el valor verdadero aún se conserva en el intervalo de sintonización (error de segundo tipo). En el primer caso, la función de pérdidas se asume igual  $A = \text{const.}$ , y para el segundo caso,  $B = \text{const.}$

## Problemas

**Problema 2.21.** Sea la trayectoria  $\mathbf{y} = \{y_k : k = 1, \dots, n\}$ ,  $n \rightarrow \infty$ , con distribución  $\mathcal{N}(\theta, 1)$ , sobre la cual se estima el parámetro  $\theta$ , para una función de riesgo  $f_c(\varepsilon) = (\theta - \varepsilon)^2$ . Calcular la función de pérdida para los siguientes casos de función de decisión Bayesiana:  $D_1 = m_{1\theta}$ ,  $D_2 = m_e(\theta)$ ,  $D_3 \equiv 0$ .

**Problema 2.22.** Resolver el problema anterior empleando los métodos de máxima verosimilitud y minimax para la selección de la mejor función de decisión.

**Problema 2.23.** Sea la observación  $\mathbf{y} = \{y_k : k = 1, \dots, n\}$  con FDP Gaussiana  $\mathcal{N}(\theta, \sqrt{\theta})$ ,  $\theta > 0$ . Estimar el parámetro  $\theta$  empleando la función de pérdida  $f_c(\varepsilon)$  dada en (2.80a) [20].



# Capítulo 3

## Representación de señales aleatorias

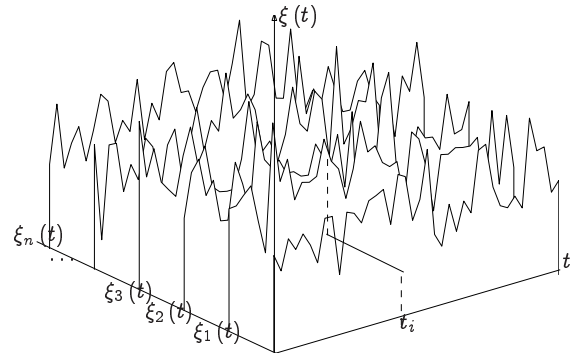
La *Teoría de procesos estocásticos* estudia las variables aleatorias, que dependen de un parámetro susceptible de cambiar de valor

B.R. Levin

La señal aleatoria  $\xi(s)$  corresponde a un proceso, que se desarrolla sobre la variable  $s$ . Cuando la variable del argumento de la señal aleatoria, que usualmente es el tiempo, discurre continuamente, entonces se habla de *funciones aleatorias*. Si en cambio, la variable del tiempo corresponde a una malla de valores discretos,  $s_1, s_2, \dots, s_n$ , entonces se habla de sucesiones o *series aleatorias*. El análisis de las señales aleatorias, en gran medida, depende de si cambia o no su estructura de aleatoriedad con respecto a su variable del argumento.

### 3.1. Señales aleatorias en el tiempo

Los fenómenos, que se desarrollan a lo largo del argumento correspondiente al tiempo y descritos por una señal aleatoria  $\xi(t)$ , se pueden analizar a partir de un conjunto longitudinal de mediciones o registros de valores  $\{\xi_i(t) \in \xi : i = 1, \dots, N\}$  como se ilustra en la Figura 3.1. En este caso, se define como *trayectoria* u observación a cada una de las mediciones simultáneas  $\xi_i$  de un misma señal aleatoria. Se define como *ensamble* o conjunto a todas las posibles trayectorias,  $\xi_i(t)$ , medidas o registradas en un intervalo de observación, que se relacionan con un mismo fenómeno aleatorio,  $\xi(t)$ .



**Figura 3.1.** Trayectorias de una señal aleatoria

En cualquier caso, las señales aleatorias pueden ser descritas mediante un ensamble de múltiples observaciones, entonces, se generan dos clases diferentes de promedios: se pueden efectuar mediciones sucesivas a lo largo de una misma observación  $\xi(t)$  a partir de las cuales se hallan sus momentos y valores de aleatoriedad o *valores promedios de tiempo*  $E\{\xi_i^n(t)\}$ , así mismo, se pueden examinar todas las observaciones del ensamble, en un momento de tiempo dado  $t_i$  con lo cual se hallan los *valores promedios de ensamble*  $E\{\xi^n(t_i)\}$ .

### 3.1.1. Estacionariedad de las señales aleatorias

Las características básicas de la estructura de aleatoriedad de las señales corresponden a los momentos y valores de aleatoriedad descritos en el numeral §2.1, los cuales pueden cambiar en el tiempo. Así por ejemplo, la FDP con dimensión múltiple  $p(\xi_1, \xi_2, \dots, \xi_n; t_1, t_2, \dots, t_n)$  corresponde a los valores instantáneos de las respectivas FDP singulares  $\xi_n(t_n)$ .

Un proceso se considera *estacionario* cuando su estructura de aleatoriedad no cambia en el tiempo, en particular, un proceso estocástico se define como *estacionario en el sentido angosto*, si sus funciones de probabilidad no son variables en el tiempo, esto es:

$$p(\xi_i, t) = p(\xi_i, t + \Delta t), \quad \forall \Delta t$$

condición, que prácticamente es bastante difícil de comprobar.

En cambio, un proceso estocástico  $\xi(t)$  se define como *estacionario en el sentido amplio*, cuando todos sus momentos y valores de aleatoriedad no varían para cualquiera que sean los tiempos de análisis  $(t_1, t_2, \dots)$ .

$$\begin{aligned} E\{\xi^n(t_n)\} &= \text{const.} \\ E\{(\xi(t_m) - m_{1\xi}(t_m))(\xi(t_n) - m_{1\xi}(t_n))\} &= \text{const} \end{aligned}$$

En consecuencia, los momentos (2.6), (2.9), (2.44) y (2.45) de las señales estacionarias se asocian con los siguientes *promedios de tiempo*:

- *Valores medios*

$$E\{\xi^n(t)\} = \int_{-\infty}^{\infty} \xi^n(t) dt \triangleq \overline{\xi^n(t)}, \quad n \in \mathbb{N} \quad (3.1)$$

- *Valores medios centralizados*

$$E\{\xi^n(t) - \overline{\xi(t)}\} = \int_{-\infty}^{\infty} (\xi(t) - \overline{\xi(t)})^n dt, \quad n \geq 2, n \in \mathbb{N} \quad (3.2)$$

- Valores de correlación

$$\begin{aligned}
 K_{\xi\eta}(t_m, t_n) &= E \{ (\xi(t_m) - m_{1\xi}(t_m)) (\eta(t_n) - m_{1\eta}(t_n)) \} \\
 &= \int_{-\infty}^{\infty} (\xi(t_m) - m_{1\xi}(t_m)) (\eta^*(t_n) - m_{1\eta}(t_n)) dt \\
 &= R_{\xi\eta}(t_m, t_n) - m_{1\xi}(t_m) m_{1\eta}(t_n)
 \end{aligned} \tag{3.3a}$$

$$R_{\xi\eta}(t_m, t_n) = E \{ \xi(t_m) \eta^*(t_n) \} = \int_{-\infty}^{\infty} \xi(t_m) \eta^*(t_n) dt \tag{3.3b}$$

Cuando en los núcleos de valores de correlación se tiene el caso de dos variables aleatorias diferentes,  $\xi(t) \neq \eta(t)$ , se habla de la *función de correlación mutua*, mientras en el caso de análisis de una misma función,  $\xi(t) = \eta(t)$ , se dice de *función de correlación propia*. Si se tienen dos variables aleatorias  $\xi(t)$  y  $\eta(t)$  con los respectivos valores de correlación propia  $R_{\xi}(t_m, t_n)$  y  $R_{\eta}(t_m, t_n)$ , en analogía con la matriz (2.54), de forma adicional se considera la *matriz de correlación*:

$$\mathbf{R} = \begin{bmatrix} R_{\xi}(t_m, t_n) & R_{\xi\eta}(t_m, t_n) \\ R_{\eta\xi}(t_m, t_n) & R_{\eta}(t_m, t_n) \end{bmatrix}$$

**Ejemplo 3.1.** Sea la suma de dos señales aleatorias

$$\eta(t) = \xi_1(t) + \xi_2(t)$$

El valor medio del proceso resultante se determina como:

$$\begin{aligned}
 E \{ \eta(t) \} &= E \{ \xi_1(t) + \xi_2(t) \} = E \{ \xi_1(t) \} + E \{ \xi_2(t) \} \\
 &= m_{1\xi_1}(t) + m_{1\xi_2}(t)
 \end{aligned}$$

Mientras, la función de correlación de la señal aleatoria resultante será:

$$R_{\eta}(t_m, t_n) = E \{ \xi_1(t_m) - m_{1\xi_1}(t_m) \} E \{ \xi_2(t_n) - m_{1\xi_2}(t_n) \}$$

que se calcula como

$$\begin{aligned}
 &= E \{ \xi_1(t_m) - m_{1\xi_1}(t_m) + \xi_2(t_m) - m_{1\xi_2}(t_m) \} E \{ \xi_1(t_n) - m_{1\xi_1}(t_n) + \xi_2(t_n) - m_{1\xi_2}(t_n) \} \\
 &= E \{ \xi_1(t_m) - m_{1\xi_1}(t_m) \} E \{ \xi_1(t_n) - m_{1\xi_1}(t_n) \} + E \{ \xi_1(t_m) - m_{1\xi_1}(t_m) \} \\
 &\quad \times E \{ \xi_2(t_n) - m_{1\xi_2}(t_n) \} + E \{ \xi_2(t_m) - m_{1\xi_2}(t_m) \} E \{ \xi_1(t_n) - m_{1\xi_1}(t_n) \} \\
 &\quad + E \{ \xi_2(t_m) - m_{1\xi_2}(t_m) \} E \{ \xi_2(t_n) - m_{1\xi_2}(t_n) \}
 \end{aligned}$$

con lo cual se tiene que

$$R_{\eta}(t_m, t_n) = R_{\xi_1}(t_m, t_n) + R_{\xi_1\xi_2}(t_m, t_n) + R_{\xi_2\xi_1}(t_m, t_n) + R_{\xi_2}(t_m, t_n)$$

Al considerar que las variables  $\xi_1(t)$  y  $\xi_2(t)$  son no correlacionadas, entonces

$$R_{\eta}(t_m, t_n) = R_{\xi_1}(t_m, t_n) + R_{\xi_2}(t_m, t_n)$$

Asumiendo que se analiza la suma de la señal aleatoria  $\xi_1(t)$  con otra señal no aleatoria  $x(t)$ , se obtiene el siguiente valor medio:

$$\begin{aligned} E\{\eta(t)\} &= E\{\xi_1(t) + x(t)\} \\ &= E\{\xi_1(t)\} + x(t) \end{aligned}$$

con la respectiva función de correlación

$$R_\eta(t_m, t_n) = R_{\xi_1}(t_m, t_n)$$

Si la variable  $x(t)$  se convierte en constante  $x(t) = c$ , se obtienen los momentos:

$$\begin{aligned} E\{\eta(t)\} &= E\{\xi_1(t) + c\} = m_{\xi_1}(t) + m_{1c} \\ R_\eta(t_m, t_n) &= R_{\xi_1}(t_m, t_n) + \sigma_c^2 \end{aligned}$$

Se puede demostrar, que para la suma  $\eta(t) = a\xi_1(t) + b\xi_2(t)$  se tienen los momentos:

$$\begin{aligned} E\{\eta(t)\} &= am_{1\xi_1}(t) + bm_{1\xi_2}(t) \\ R_\eta(t_m, t_n) &= a^2 R_{\xi_1}(t_m, t_n) + b^2 R_{\xi_2}(t_m, t_n) + ab(R_{\xi_1\xi_2}(t_m, t_n) + R_{\xi_2\xi_1}(t_m, t_n)) \end{aligned}$$

**Ejemplo 3.2.** Hallar la función de correlación de la siguiente suma de señales aleatorias

$$\xi(t) = \sum_{n=1}^N \xi_n(t)$$

donde  $\xi_n(t) = \xi_{nc} \cos \omega_n t + \xi_{ns} \sin \omega_n t$ , siendo  $\xi_c$  y  $\xi_s$  valores aleatorios no correlacionados con media cero e igual varianza  $\sigma_{\xi_{nc}}^2 = \sigma_{\xi_{ns}}^2 = \sigma_{\xi_n}^2$ .

La función de correlación de  $\xi_n(t) = \xi_{nc} \cos \omega_n t + \xi_{ns} \sin \omega_n t$  se calcula como:

$$\begin{aligned} R_{\xi_n}(t_1, t_2) &= E\{\xi_n(t_1)\xi_n(t_2)\} \\ &= E\{(\xi_{nc} \cos \omega_n t_1 + \xi_{ns} \sin \omega_n t_1)(\xi_{nc} \cos \omega_n t_2 + \xi_{ns} \sin \omega_n t_2)\} \\ &= \sigma_{\xi_{nc}}^2 \cos \omega_n t_1 \cos \omega_n t_2 + \sigma_{\xi_{ns}}^2 \sin \omega_n t_1 \sin \omega_n t_2 \\ &= \sigma_{\xi_n}^2 \cos \omega_n (t_1 - t_2) \end{aligned}$$

que da como resultado

$$R_\xi(t_1, t_2) = \sum_{n=1}^N R_{\xi_n}(t_1, t_2) = \sum_{n=1}^N \sigma_{\xi_n}^2 \cos \omega_n (t_1 - t_2)$$

**Transformación lineal de señales aleatorias.** Sea un sistema con entrada  $x$  y salida  $y$ , descrito por un operador lineal  $\mathcal{K}$ , tal que cumpla las siguientes condiciones:

1.  $y = \mathcal{K}\{\lambda x\} = \lambda \mathcal{K}\{x\}$ ,  $\lambda = \text{const.}$
2.  $y = \mathcal{K}\{x_1 + x_2\} = \mathcal{K}\{x_2\} + \mathcal{K}\{x_1\}$

En los casos de transformación de variables, son importantes los valores de aleatoriedad y momentos de los procesos de entrada o salida, dada la descripción del sistema en la forma  $\eta = \mathcal{K}\{\xi\}$ .

En particular, el valor medio y la función de correlación de la salida son de la forma:

$$m_{1\eta}(t) = E\{\eta(t)\} = E\{\mathcal{K}\{\xi(t)\}\} = \mathcal{K}\{E\{\xi(t)\}\} \quad (3.4a)$$

$$\begin{aligned} R_{\eta}(t_m, t_n) &= E\{\eta(t_m) - m_{1\eta}(t_m)\} E\{\eta(t_n) - m_{1\eta}(t_n)\} \\ &= E\{\mathcal{K}\{\xi(t_m) - m_{1\xi}(t_m)\}\} E\{\mathcal{K}\{\xi(t_n) - m_{1\xi}(t_n)\}\} \\ &= \mathcal{K}\{E\{\xi(t_m) - m_{1\xi}(t_m)\}\} \mathcal{K}\{E\{\xi(t_n) - m_{1\xi}(t_n)\}\} \\ &= \mathcal{K}_{(t_m)} \mathcal{K}_{(t_n)} \{E\{\xi(t_m) - m_{1\xi}(t_m)\} E\{\xi(t_n) - m_{1\xi}(t_n)\}\} \\ &= \mathcal{K}_{(t_m)} \mathcal{K}_{(t_n)} \{R_{\xi}(t_m, t_n)\} \end{aligned} \quad (3.4b)$$

Cabe anotar que la suposición de estacionariedad simplifica las expresiones (3.4a) y (3.4b), obtenidas para la transformación de señales, así:

$$E\{\eta(t)\} = E\{\mathcal{K}\{\xi(t)\}\} = \mathcal{K}\{E\{\xi(t)\}\} = m_{1\eta} \quad (3.5a)$$

$$\mathcal{K}_{(t_m)} \mathcal{K}_{(t_n)} \{R_{\xi}(t_m, t_n)\} = \mathcal{K} \mathcal{K}_{(\tau)} \{R_{\xi}(\tau)\}, \quad \tau = t_m - t_n \quad (3.5b)$$

De otra parte, la reacción de un sistema lineal e invariante en el tiempo (1.39), de acuerdo con (1.33), se describe por la reacción del operador lineal del sistema a la función  $\delta(t)$ , o respuesta a impulso  $h(t)$ , por lo que si a la entrada del sistema se tiene la señal aleatoria  $\xi(t)$ , mientras a la salida se tiene la señal aleatoria  $\eta(t)$ , entonces, se observan las siguientes relaciones en el tiempo entre los momentos de entrada y salida del sistema:

$$m_{1\eta}(t) = \int_0^{\infty} m_{1\xi}(t - \tau) h(\tau) d\tau \quad (3.6a)$$

$$R_{\eta}(t_m, t_n) = \int_0^{\infty} \int_0^{\infty} R_{\xi}(t_m - \tau_m, t_n - \tau_n) h(\tau_m) h(\tau_n) d\tau_m d\tau_n \quad (3.6b)$$

**Ejemplo 3.3.** Dados, para la variable  $\xi(t)$ , la media  $m_{1\xi}(t) = at^2$  y la función de correlación

$$R_{\xi}(t_m, t_n) = \sigma_{\xi}^2 \exp -\alpha |t_m - t_n|$$

Hallar los momentos  $m_{1\eta}$ ,  $R_{\eta}(t_m, t_n)$  y varianza de la señal aleatoria en los siguientes dos casos de análisis de transformación lineal:

$$a) \eta(t) = \frac{d\xi}{dt}, \quad b) \eta(t) = \int_0^t \xi(t) dt$$

En el caso a), el valor medio es igual a:

$$\begin{aligned} m_{1\eta}(t) &= \mathcal{K}\{E\{\xi(t)\}\} = \frac{d}{dt} m_{1\xi}(t) \\ &= 2at \end{aligned}$$

La función de correlación del proceso de salida se halla diferenciando dos veces la función de

correlación del proceso de entrada, de la siguiente manera:

$$\begin{aligned} t_m > t_n : R_\xi(t_m, t_n) &= \sigma_\xi^2 e^{-\alpha(t_m - t_n)} \\ \frac{\partial}{\partial t_m} R_\xi(t_m, t_n) &= -\sigma_\xi^2 \alpha e^{-\alpha(t_m - t_n)} \\ \frac{\partial^2}{\partial t_m \partial t_n} R_\xi(t_m, t_n) &= \sigma_\xi^2 \alpha^2 e^{-\alpha(t_m - t_n)} \end{aligned}$$

$$\begin{aligned} t_m < t_n : R_\xi(t_m, t_n) &= \sigma_\xi^2 e^{-\alpha(t_n - t_m)} \\ \frac{\partial}{\partial t_m} R_\xi(t_m, t_n) &= \sigma_\xi^2 \alpha e^{-\alpha(t_n - t_m)} \\ \frac{\partial^2}{\partial t_m \partial t_n} R_\xi(t_m, t_n) &= -\sigma_\xi^2 \alpha^2 e^{-\alpha(t_n - t_m)} \end{aligned}$$

Juntar ambos casos de análisis ( $t_m > t_n$  y  $t_m < t_n$ ) es imposible. Por lo tanto, se puede inferir que la función de correlación  $R_\xi(t_m, t_n)$  no tiene la segunda derivada para el valor  $t_m = t_n$ , luego, el proceso  $\xi(t)$  no es diferenciable.

En el caso b), la media  $m_{1\xi}(t) = at^3/3$ . Mientras, la función de correlación se define como

$$R_\xi(t_m, t_n) = \sigma_\xi^2 \int_0^{t_m} \int_0^{t_n} e^{-\alpha|\lambda_m - \lambda_n|} d\lambda_m d\lambda_n$$

Luego, se tienen los siguientes casos de análisis:

$$\begin{aligned} R_\xi(t_m, t_n) &= \sigma_\xi^2 \int_0^{t_m} \int_0^{t_n} e^{-\alpha|\lambda_m - \lambda_n|} d\lambda_n d\lambda_m \\ t_n > t_m : \int_0^{t_n} e^{-\alpha|\lambda_m - \lambda_n|} d\lambda_n &= \int_0^{\lambda_m} e^{-\alpha(\lambda_m - \lambda_n)} d\lambda_n + \int_{\lambda_m}^{t_n} e^{-\alpha(\lambda_n - \lambda_m)} d\lambda_n \\ &= \frac{1}{\alpha} e^{-\alpha(\lambda_m - \lambda_n)} \Big|_0^{\lambda_m} - \frac{1}{\alpha} e^{\alpha(\lambda_m - \lambda_n)} \Big|_{\lambda_m}^{t_n} = \frac{1}{\alpha} \left( 2 - e^{-\alpha\lambda_m} - e^{-\alpha(t_n - \lambda_m)} \right); \\ &\int_0^{t_m} \frac{1}{\alpha} \left( 2 - e^{-\alpha\lambda_m} - e^{-\alpha(t_n - \lambda_m)} \right) d\lambda_m \\ &= \frac{1}{\alpha} \left( 2\lambda_m + \frac{1}{\alpha} e^{-\alpha\lambda_m} - \frac{1}{\alpha} e^{-\alpha(t_n - \lambda_m)} \right) \Big|_0^{t_m} \\ &= \frac{1}{\alpha^2} \left( 2\alpha t_m - 1 + e^{-\alpha t_m} + e^{-\alpha t_n} - e^{-\alpha(t_n - t_m)} \right) \\ R_\eta(t_m, t_n) &= \frac{\sigma_\xi^2}{\alpha^2} \left( 2\alpha t_m - 1 + e^{-\alpha t_m} + e^{-\alpha t_n} - e^{-\alpha(t_n - t_m)} \right) \end{aligned}$$

El análisis de  $t_m > t_n$ , muestra que de manera similar se obtiene:

$$R_\eta(t_m, t_n) = \frac{\sigma_\xi^2}{\alpha^2} \left( 2\alpha t_n - 1 + e^{-\alpha t_m} + e^{-\alpha t_n} - e^{-\alpha(t_m - t_n)} \right)$$

Al juntar ambas expresiones obtenidas en una sola función, que generalice para todo valor de  $t_m$  y  $t_n$ , se tiene:

$$R_\eta(t_m, t_n) = \frac{\sigma_\xi^2}{\alpha^2} \left( 2\alpha \min\{t_m, t_n\} - 1 + e^{-\alpha t_m} + e^{-\alpha t_n} - e^{-\alpha|t_m - t_n|} \right)$$

Finalmente, basados en lo anterior se puede encontrar la varianza, haciendo  $t_m = t_n = t$

$$\sigma_\eta^2(t) = 2\sigma_\xi^2 (\alpha t - 1 + e^{-\alpha t}) / \alpha^2$$

**Ejemplo 3.4.** Un proceso aleatorio corresponde a la transformación  $\xi(t) = \phi f(t)$ , siendo  $f$  una función determinística y  $\phi$  una variable aleatoria. Determinar si el proceso  $\xi(t)$  es estacionario.

La estacionariedad en el sentido amplio, exige la invariabilidad en el tiempo de los valores de aleatoriedad, por ejemplo del primer momento inicial (3.1) y la varianza (3.2),

$$\begin{aligned} E\{\xi(t)\} &= E\{\phi f(t)\} = f(t) E\{\phi\} \\ E\left\{\left(\xi(t) - \overline{\xi(t)}\right)^2\right\} &= E\left\{\left(\phi f(t) - \overline{\phi f(t)}\right)^2\right\} \\ &= f^2(t) E\{\phi\} \end{aligned}$$

que resultan variantes en el tiempo, y por ende, el proceso  $\xi(t)$  no es estacionario.

En la práctica, es frecuente la remoción del valor medio, y la consecuente normalización del proceso aleatorio  $\xi(t)$  sobre la desviación estándar,

$$\xi_s(t) = \frac{\xi(t) - \overline{\xi(t)}}{\sigma_\xi}$$

Sin embargo, hay que tener en cuenta que la FDP del proceso estandarizado es diferente de la densidad de probabilidad inicial, en particular, teniendo en cuenta la expresión para la transformación de variables (2.34), se obtiene

$$p(\xi) = \frac{1}{\sigma_\xi} p\left(\sigma_\xi \xi_s + \overline{\xi(t)}\right)$$

### 3.1.2. Ergodicidad de las señales aleatorias

Sea el proceso estacionario, en el cual se conocen los valores del primer momento inicial de cada una de las  $n$  trayectorias. Considerando la estacionariedad del proceso, la esperanza matemática del mismo en cualquier corte de tiempo  $t_i$ , como se muestra en la Figura 3.1, se determina de la forma

$$m_{1\xi(t_i)} = \frac{1}{n} \sum_{k=1}^n \xi_k(t_i)$$

El cálculo de primer momento, por alguna de las observaciones definida en un intervalo de análisis  $T$ , se puede realizar segmentando el intervalo en una malla de subintervalos,

determinados sobre los momentos de tiempo  $t_k = k\Delta t$ ,  $k = 1, \dots, n$ , esto es,  $n = T/\Delta t$ . La estimación de la esperanza sobre la malla de subintervalos se determina de la expresión:

$$\tilde{m}_{1\xi_i} = \frac{1}{n} \sum_{k=1}^n \xi_i(t_k) \Delta t$$

que al hallar el límite, resulta en

$$\tilde{m}_{1\xi_i} \underset{n \rightarrow \infty}{\approx} \frac{1}{T} \int_0^T \xi_i(t) dt$$

Por cuanto, se asume que las propiedades de aleatoriedad de los procesos estacionarios se mantienen invariantes en el tiempo, entonces, cuando se tiene un número  $n$  suficientemente grande, para cada trayectoria se cumple la igualdad

$$\tilde{m}_{1\xi_i} = m_{1\xi(t_i)}, \quad i = 1, \dots, n$$

Generalizando la expresión anterior, un proceso aleatorio estacionario se define como *ergódico*, cuando los valores promedios de tiempo y de ensamble son idénticos, esto es,

$$E \{ \xi_i^n(t) \} = E \{ \xi^n(t_i) \} \quad (3.7)$$

La media de cada observación del ensamble es igual a la media de la observación media del ensamble. De (3.7), se entiende que un proceso ergódico es estacionario, pero no viceversa.

La estimación de los momentos (3.1) para una señal ergódica  $\xi(t)$  se describe como

$$\tilde{E} \{ \xi^n(t) \} = \int_{-\infty}^{\infty} \xi^n(t) w(t) dt, \quad n \in \mathbb{Z} \quad (3.8)$$

siendo  $w(t)$  la función de peso, que cumple la restricción

$$\lim_{T \rightarrow \infty} \int_0^T w(t) dt = 1 \quad (3.9)$$

La función de peso rectangular es la más frecuentemente empleada en la estimación de los momentos en (3.8),

$$w(t) = \begin{cases} 1/T, & 0 \leq t \leq T \\ 0, & \text{otros valores de } t \end{cases} \quad (3.10)$$

La función de peso (3.10) corresponde a la respuesta a impulso (1.35) de un filtro lineal,



cuya respectiva función de transferencia es denominada *característica espectral*.

$$\Phi(\omega) = \mathcal{F}\{w(t)\} = \int_{-\infty}^{\infty} w(t) e^{-j\omega t} dt$$

que en el caso particular de (3.10) es igual a

$$\Phi(\omega) = e^{-j\omega T/2} \text{sinc}(\omega T/2)$$

Los valores de aleatoriedad dados en (3.1), (3.2), (3.3a) y (3.3b) para los procesos ergódicos, teniendo en cuenta la estimación (3.10), toman las respectivas expresiones:

$$E\{\xi^n(t)\} = \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T \xi^n(t) dt, \quad n \in \mathbb{N} \quad (3.11a)$$

$$E\{\xi^n(t) - \overline{\xi(t)}\} = \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T (\xi(t) - \overline{\xi(t)})^n dt, \quad n \geq 2, n \in \mathbb{N} \quad (3.11b)$$

$$R_\xi(\tau) = \lim_{T \rightarrow \infty} \frac{1}{T} \int_T (\xi(t))(\xi^*(t+\tau)) dt \quad (3.11c)$$

$$\begin{aligned} K_\xi(\tau) &= \lim_{T \rightarrow \infty} \frac{1}{T} \int_T (\xi(t) - \overline{\xi(t)}) (\xi^*(t+\tau) - \overline{\xi(t)}) dt \\ &= R_\xi(\tau) - m_{1\xi}^2 \end{aligned} \quad (3.11d)$$

Los momentos de las señales ergódicas se relacionan con el siguiente sentido físico:

- Valor medio  $\overline{\xi(t)}$  es la *componente constante*.
- El valor medio al cuadrado  $\overline{\xi(t)^2}$  corresponde a la *potencia de la componente directa*.
- Valor cuadrático medio  $\overline{\xi^2(t)}$  es la *potencia promedio*.
- La varianza  $\sigma_\xi^2$  es la *potencia de la componente alterna*.
- La desviación estándar  $\sigma_\xi$  corresponde al *valor rms*.
- La función de correlación propia corresponde a la *potencia de la componente alterna*

**Ejemplo 3.5.** Hallar la función de correlación de la señal aleatoria

$$\xi(t) = a \cos(\omega t + \phi)$$

siendo  $\phi$  la variable aleatoria.

A partir de la definición (3.11c) se tiene que

$$\begin{aligned}
R_{\xi}(\tau) &= \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T a \cos(\omega t + \phi) a \cos(\omega t + \omega \tau + \phi) dt \\
&= \lim_{T \rightarrow \infty} \frac{a^2}{T} \int_0^T \frac{1}{2} (\cos(\omega \tau + \phi) + \cos(2\omega t + \omega \tau + 2\phi)) dt
\end{aligned}$$

La primera integral se calcula de la forma

$$\frac{a^2}{2T} \int_0^T \cos \omega \tau dt = \frac{a^2}{2} \cos \omega \tau$$

Mientras, la segunda integral es igual a

$$\begin{aligned}
\frac{a^2}{2T} \int_0^T \cos(2\omega t + 2\phi + \omega \tau) dt \\
= \frac{a^2}{2T} \cos(2\phi + \omega \tau) \int_0^T \cos 2\omega t dt - \frac{a^2}{2T} \sin(2\phi + \omega \tau) \int_0^T \sin 2\omega t dt
\end{aligned}$$

Por cuanto,

$$\lim_{T \rightarrow \infty} \int_T \cos k\omega t dt = \lim_{T \rightarrow \infty} \text{sinc}(k\omega T) = 0, \quad \lim_{T \rightarrow \infty} \int_T \sin k\omega t dt = 0, \quad \forall k \in \mathbb{Z}$$

entonces, finalmente se obtiene la siguiente expresión para la función de correlación:

$$R_{\xi}(\tau) = \frac{a^2}{2} \cos \omega \tau$$

Por su papel preponderante en la descripción de procesos estacionarios que las funciones de correlación y covarianza tienen se debe tener en cuenta sus siguientes propiedades:

(a). *Paridad.*  $R_{\xi}(\tau) = R_{\xi}^*(-\tau)$ ,  $K_{\xi}(\tau) = K_{\xi}(-\tau)$

A partir de la estacionariedad del proceso se puede demostrar la simetría de los valores de correlación respecto a los argumentos  $t_1$  y  $t_2$ :

$$E \left\{ \xi_m(t_1) - \overline{\xi_m(t_1)}, \xi_n(t_2) - \overline{\xi_n(t_2)} \right\} = E \left\{ \xi_n(t_2) - \overline{\xi_n(t_2)}, \xi_m(t_1) - \overline{\xi_m(t_1)} \right\}$$

(b). *Valor máximo.*  $|R_{\xi}(\tau)| \leq R_{\xi}(0)$ ,  $|K_{\xi}(\tau)| \leq K_{\xi}(0)$ .

Por cuanto es evidente la siguiente desigualdad  $E \left\{ (\xi(t) \pm \xi(t + \tau))^2 \right\} \geq 0$ , abriendo paréntesis, se obtiene

$$\overline{\xi^2(t)} + \overline{\xi^2(t + \tau)} \pm 2\overline{\xi(t)\xi(t + \tau)} \geq 0$$

Debido a la estacionariedad del proceso, los dos primeros términos son iguales, con lo cual se obtiene la cota máxima de la función.

$$\overline{\xi^2(t)} \geq \pm \overline{\xi(t)\xi(t+\tau)} \Rightarrow R_\xi(0) \geq R_\xi(\tau)$$

Por cierto, de (3.3a) se tienen los siguiente valores en el origen:

$$K_\xi(0) = \frac{1}{T} \int_0^T \left( \xi(t) - \overline{\xi(t)} \right) \left( \xi^*(t) - \overline{\xi(t)} \right) dt = \sigma_\xi^2, \quad (3.12a)$$

$$R_\xi(0) = \frac{1}{T} \int_0^T \xi(t) \xi^*(t) dt = \overline{\xi^2(t)} \quad (3.12b)$$

- (c). *Continuidad.* Si la función  $R_\xi(\tau)$  es continua en  $\tau = 0$ , entonces  $R_\xi(\tau)$  también es continua para todo  $\tau$ .
- (d). *Periodicidad.*  $R_\xi(\tau) = R_\xi(\tau + T), \forall \tau \in T$ .
- (e). *Restricción en la forma.* Dado un proceso físicamente realizable, la transformada de Fourier de su función de correlación debe cumplir la siguiente restricción de forma:

$$\mathcal{F}\{R_\xi(\tau)\} \geq 0, \quad \forall \omega \quad (3.13)$$

- (f). *Convergencia.* Si el proceso aleatorio  $\xi(t)$  no es periódico, entonces se cumple,

$$\lim_{|\tau| \rightarrow \infty} R_\xi(\tau) = \overline{\xi^2(t)}, \quad \lim_{|\tau| \rightarrow \infty} K_\xi(\tau) = 0,$$

En este sentido se introduce el concepto de *intervalo de correlación*, que se puede determinar, bien como la parte  $\beta \ll 1$  del valor máximo de la función de correlación, bien como la mitad de la base del rectángulo  $\Delta\tau_e$  con altura unitaria y área igual a la de la función de correlación normalizada. En el primer caso, el intervalo de correlación corresponde al valor  $\tau = \tau_c$ , tal que el valor de la función normalizada de correlación sea considerablemente pequeño:

$$\frac{R_\xi(\tau_c)}{\sigma_\xi^2} = \beta \approx 0.05 \dots 0.1 \quad (3.14)$$

mientras en el segundo caso, el intervalo se calcula a partir de la integral

$$\Delta\tau_e = \frac{1}{R_\xi(0)} \int_0^\infty R_\xi(\tau) d\tau \quad (3.15)$$

Por último, se puede decir que la función de correlación caracteriza la dependencia estadística entre los valores instantáneos del proceso, dados en diferentes momentos del tiempo.

po. Así, por ejemplo, sea la diferencia cuadrática media entre dos valores instantáneos del proceso definida de la forma,

$$\varepsilon_{\xi}(\tau) = E \left\{ (\xi(t) - \xi(t + \tau))^2 \right\}$$

de la cual se observa que

$$\begin{aligned} \varepsilon_{\xi}(\tau) &= \overline{\xi^2(t)} + \overline{\xi^2(t + \tau)} - 2\overline{\xi(t)\xi(t + \tau)} = 2\overline{\xi^2(t)} - 2R_{\xi}(\tau) \\ &= 2R_{\xi}(0) - 2R_{\xi}(\tau) \end{aligned}$$

luego,  $2R_{\xi}(\tau) + \varepsilon_{\xi}(\tau) = 2R_{\xi}(0) = \text{const.}$

Por lo tanto, la función de correlación complementa la desviación cuadrática media hasta un valor constante y proporcional a la potencia del proceso.

**Ejemplo 3.6.** Hallar el intervalo de correlación para la función de correlación de la forma,  $R_{\xi}(\tau) = k/(1 - \alpha^2\tau^2)$ .

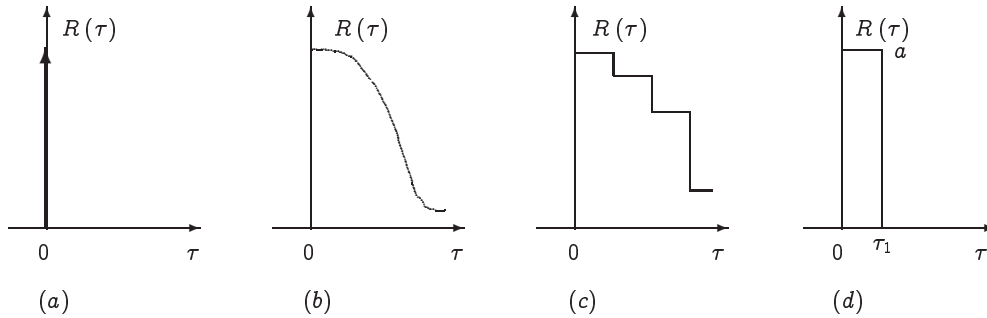
De la expresión (3.14), se tiene que

$$1/(1 - \alpha^2\tau_c^2) = \beta \Rightarrow \tau_c = \frac{1}{\alpha} \sqrt{\frac{1 - \beta}{\beta}}$$

mientras para el segundo caso (3.15) se obtiene

$$\Delta\tau = \int_0^{\infty} \frac{1}{1 + \alpha^2\tau^2} d\tau = \frac{1}{\alpha} \arctan \alpha\tau \Big|_0^{\infty} = \frac{\pi}{2\alpha}$$

Ea usual considerar las siguientes formas de clasificación de los procesos ergódicos de acuerdo a su función de correlación propia:



**Figura 3.2.** Clasificación de funciones de correlación

- Procesos con correlación nula, en los cuales el valor de la función de correlación es cero, excepto para el corte de tiempo  $R(0) \sim \delta(\tau)$  (Figura 3.2(a)).

- Procesos con correlación continua, en los cuales el valor de la función cambia sobre un dominio continuo, (Figura 3.2(b)).
- Procesos con correlación discreta o *procesos de Markov*, cuando la función de correlación cambia de forma escalonada sobre una malla discreta de valores del intervalo de correlación  $\{\tau_k\}$ , (Figura 3.2(c)). Dentro de los procesos de Markov, una clase importante corresponde al caso cuando la función de correlación discreta se limita a un solo escalón (Figura 3.2(d)) lo que se interpreta como la dependencia de dos valores contiguos del proceso.

Realmente el proceso ilustrado en la Figura 3.2(d) es un modelo abstracto, en la medida en que la función de correlación no cumple con la condición de existencia (3.13), para todos los valores del intervalo de correlación  $\tau$ , como se observa de la relación:

$$\int_0^{\infty} R(\tau) \cos \omega \tau d\tau = \int_0^{\tau_1} a \cos \omega \tau d\tau = \frac{a}{\omega} \sin \omega \tau_1 \neq 0, \quad \forall \tau$$

### 3.1.3. Descomposición espectral de señales aleatorias

Del ejemplo 3.2, se observa que la función de correlación resultante  $R_{\xi}(t_1, t_2)$  es estacionaria, en la medida en que el argumento del coseno depende sólo de la distancia entre los valores de tiempo  $|t_2 - t_1|$ . Por lo que la función de correlación resultante es:

$$R_{\xi}(t_1, t_2) = R_{\xi}(\tau) = \sum_n \sigma_{\xi_n}^2 \cos \omega_n \tau, \quad \sigma_{\xi}^2 = \sum_n \sigma_{\xi_n}^2$$

Lo anterior, plantea la tarea de representación de señales en forma de la serie generalizada de Fourier (1.4), pero ampliada al caso de procesos aleatorios estacionarios.

Sea  $\xi(t)$  un proceso estacionario en el sentido amplio, dado en el intervalo  $(0, T)$ . Entonces, su descomposición ortogonal tendrá la forma:

$$\xi(t) = \sum_{n=0}^{\infty} \xi_n \phi_n(t) \quad (3.16)$$

donde los coeficientes de descomposición, en concordancia con (1.7), se determinan como:

$$\xi_n = \frac{1}{\mathcal{E}_n} \int_0^T \xi(t) \phi_n^*(t) dt, \quad \mathcal{E}_i = \int_0^T \phi_n^2(t) dt \quad (3.17)$$

Sin embargo, a diferencia de la representación (1.4), los coeficientes de (3.17) son aleatorios, por lo que la convergencia de la suma en (3.16) hacia la señal aleatoria  $\xi(t)$ , se debe entender en el sentido del valor cuadrático medio:

$$\lim_{N \rightarrow \infty} E \left\{ \left| \xi(t) - \sum_{n=0}^{\infty} \xi_n \phi_n(t) \right|^2 \right\} = 0$$

Uno de los problemas importantes a resolver es la selección del conjunto base de representación  $\{\phi_n(t)\}$ , el cual puede ser escogido por los dos siguientes principios: Primero, que se brinde el menor error de representación para un número  $N$  dado de coeficientes, o bien, segundo, que se genere el menor número  $N$  de coeficientes para un valor dado de error de representación. Así mismo, es importante que los coeficientes (3.17) tengan correlación cero entre ellos, por cuanto las tareas asociadas al proceso de señales aleatorias se resuelven de manera más fácil.

Se puede demostrar que para una señal aleatoria, representada por la serie (3.16), que tenga función de correlación continua  $R(t, \tau)$ , el valor de la esperanza del error cuadrático medio integral (potencia media de error), expresado como:

$$E \left\{ \int_0^T \left[ \xi(t) - \sum_{n=0}^{N-1} \xi_n \phi_n(t) \right]^2 dt \right\} \quad (3.18)$$

el error será el mínimo posible (3.18), para cualquier  $N$ , si el conjunto base de representación  $\{\phi_n(t) : n = 0, 1, \dots, N-1\}$  cumple la condición homogénea de Fredholm de segundo tipo [15, 22]:

$$\lambda_i \phi_i(t) = \frac{1}{T} \int_0^T R_\xi(t, \tau) g(\tau) d\tau \quad (3.19)$$

donde  $\{\phi_n(t)\}$  son las funciones propias (solución) de la ecuación y los coeficientes  $\lambda_n$  son los valores propios del núcleo  $R_\xi(t, \tau)$  de la ecuación. El conjunto de funciones propias  $\{\phi_n(t)\}$  es ortogonal, cuyos coeficientes pueden ser escogidos de tal manera que puedan ser ortonormales. Los coeficientes de descomposición (3.16) de la señal aleatoria, en caso de emplear el conjunto base de representación  $\{\phi_n(t)\}$  que cumplan con (3.19), resultan tener correlación nula. Además, si la señal aleatoria se asume del tipo Gaussiano, entonces los coeficientes resultan ser independientes estadísticamente. Así mismo, si se cumple que  $E\{\xi_n\} = 0$ , entonces, el valor de la varianza  $\sigma_{\xi_n}^2$  converge al respectivo valor propio  $\lambda_n$ .

Para el conjunto base de representación  $\{\phi_n(t)\}$ , que cumple la condición (3.19), la esperanza de la potencia de error (3.18), dada en el intervalo  $(0, T)$ , se determina como:

$$\begin{aligned} \bar{\varepsilon}_T^2 &= \frac{1}{T} E \left\{ \int_0^T \left\{ \xi(t) - \sum_{n=0}^{N-1} \xi_n \phi_n(t) \right\}^2 dt \right\} \\ &= \frac{1}{T} \int_0^T E \{ \xi^2(t) \} dt - \frac{2}{T} \int_0^T \sum_{n=0}^{N-1} E \{ \xi_n \xi^2(t) \} \phi_n(t) dt \\ &\quad + \frac{1}{T} \int_0^T \sum_{m=0}^{N-1} \sum_{n=0}^{N-1} E \{ \xi_m \xi_n \} \phi_m(t) \phi_n(t) dt \end{aligned}$$

Por lo que se tiene que

$$\bar{\varepsilon}_T^2 = \sigma_\xi^2 - \sum_{n=0}^{N-1} \sigma_{\xi_n}^2 \quad (3.20)$$

La expresión (3.20) permite hallar el número  $N$  de componentes de la serie de descomposición (3.16), que brinde un valor a priori dado de error de representación.

La descomposición de señales aleatorias, que tengan función de correlación continua, por la serie (3.16), en la cual el conjunto base de representación corresponde a las funciones propias, se denomina *descomposición de Karhunen-Loève* (K-L).

El ruido blanco Gaussiano, mostrado en el literal §3.1.4 con función de correlación (3.34), se puede descomponer en el intervalo  $(0, T)$  empleando cualquier conjunto base ortogonal, pero en todo caso, los coeficientes de descomposición serán valores aleatorios Gaussianos estadísticamente independientes con igual varianza de valor  $N_0/2$ .

**Ejemplo 3.7.** Sean  $\xi(t)$  y  $\eta(t)$ , señales aleatorias dadas en un intervalo finito de tiempo  $t \in T_\xi$ . Hallar la función de correlación mutua usando la descomposición ortogonal de Fourier. Las series de Fourier (1.8), para las señales aleatorias,  $\xi(t)$  y  $\eta(t)$ , tienen la misma forma,

$$x(t) = a_0 + \sum_{n=1}^{\infty} (a_n \cos n\Omega t + b_n \sin n\Omega t), \quad \Omega = \frac{2\pi}{T_\xi}, \quad x \in \{\xi, \eta\}$$

La descomposición de Fourier de la versión con desfase de cada proceso tiene la forma,

$$\begin{aligned} x(t + \tau) = & a_0 + \sum_{n=1}^{\infty} a_n \cos n\Omega \tau \cos n\Omega t - \sum_{n=1}^{\infty} a_n \sin n\Omega \tau \sin n\Omega t + \\ & + \sum_{n=1}^{\infty} b_n \sin n\Omega \tau \cos n\Omega t + \sum_{n=1}^{\infty} b_n \cos n\Omega \tau \sin n\Omega t \end{aligned}$$

Reemplazando ambas expresiones, en la definición de la función de correlación mutua,

$$R_{\xi\eta}(\tau) = a_{0\xi}a_{0\eta} + \frac{1}{2} \sum_{n=1}^{\infty} (a_{n\xi}a_{n\eta} + b_{n\xi}b_{n\eta}) \cos n\Omega \tau + \frac{1}{2} \sum_{n=1}^{\infty} (a_{n\xi}b_{n\eta} - a_{n\eta}b_{n\xi}) \sin n\Omega \tau$$

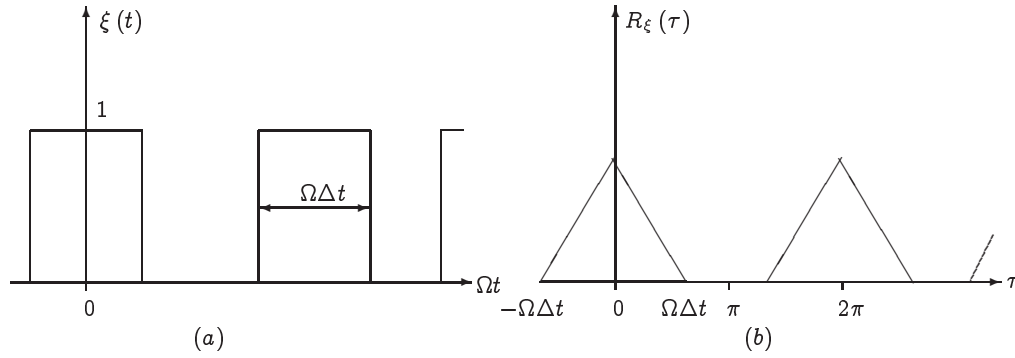
Cuando se asume la paridad de ambas señales aleatorias en análisis, entonces, los términos del senos se convierten en 0;  $b_\xi = b_\eta = 0$ , luego, la anterior expresión se simplifica hasta,

$$R_{\xi\eta}(\tau) = a_{0\xi}a_{0\eta} + \frac{1}{2} \sum_{n=1}^{\infty} a_{n\xi}a_{n\eta} \cos n\Omega \tau$$

**Ejemplo 3.8.** Hallar la función de correlación propia para el tren de pulsos cuadrados, mostrado en la Figura 3.3(a), empleando la descomposición ortogonal de Fourier.

La serie de Fourier de la función tren de pulsos cuadrados tiene la forma

$$\xi(t) = \frac{\Omega \Delta t}{2} + \frac{2}{\pi} \sum_{n=1}^{\infty} \frac{1}{n} \sin n \frac{\Omega \Delta t}{n} \cos n\Omega t$$



**Figura 3.3.** Función periódica pulso cuadrado

la cual se reemplaza por la expresión, obtenida en el ejemplo anterior para el caso de funciones pares, haciendo  $a_\xi = a_\eta$ , con lo que la función de correlación propia resulta en

$$\begin{aligned} R_\xi(\tau) &= a_{0\xi}^2 + \frac{1}{2} \sum_{n=1}^{\infty} a_{n\xi}^2 \cos n\Omega\tau \\ &= \left( \frac{\Omega\Delta t}{2\pi} \right)^2 + \frac{4}{\pi^2} \frac{1}{2} \sum_{n=1}^{\infty} \frac{1}{n^2} \sin^2 n \frac{\Omega\Delta t}{2} \cos n\Omega\tau \end{aligned}$$

Teniendo en cuenta la siguiente expresión cerrada, obtenida para la suma [13]:

$$\sum_{n=1}^{\infty} \sin^2 \frac{n\alpha}{n^2} \cos nx = \begin{cases} -\frac{\pi}{4}x + \frac{\pi\alpha}{2} - \frac{\alpha^2}{2}, & 0 < x < 2\alpha \\ -\frac{\alpha^2}{2}, & 2\alpha < x < \pi \end{cases}$$

entonces, se observa claramente, que la función de correlación corresponde a un tren de pulsos triangulares como se muestra en la Figura 3.3(b).

### 3.1.4. Densidad espectral de potencia

La descripción directa de una señal aleatoria  $\xi(t)$  en el dominio espectral, mediante la TF,

$$\Xi(\omega) = \int_{-\infty}^{\infty} \xi(t) e^{-j\omega t} dt$$

es imposible de realizar, principalmente, porque la integral (1.18) implica cumplir las condiciones de Dirichlet (sección §1.2.1), las cuales no ocurren para cualquier observación de proceso estacionario en el sentido amplio, particularmente, la condición de convergencia exige que el proceso sea absolutamente integrable; condición que sólo se alcanzaría para señales aleatorias de contenido cero.

El empleo de la TF exige la modificación de la representación para las observaciones de la señal estacionaria, de tal manera que la integral (1.18) converja, para lo cual, la forma



más sencilla consiste en el truncamiento de la trayectoria  $\xi(t)$ , empleado en el cálculo de sus momentos de una observación sobre un intervalo de tiempo  $(-T, T)$ , descrito en (3.11a) y (3.11b), que usa la función ventana rectangular:

$$\xi_T(t) = \text{rect}_T(t) \xi(t) \quad (3.21)$$

La trayectoria de la señal aleatoria truncada de (3.21) es válida, mientras la varianza del proceso sea finita. Así, se asegura la convergencia de la integral (1.18).

De otra parte, el empleo del teorema de Parsevall, mostrado en la Tabla 1.2, para la observación truncada (3.21), asumiendo su valor medio igual a cero, da como resultado:

$$\int_{-T}^T \xi_T^2(t) dt = \frac{1}{2\pi} \int_{-\infty}^{\infty} |\Xi_T(\omega)|^2 d\omega$$

cuyo promedio se obtiene acorde con la ventana de estimación (3.10):

$$\frac{1}{2T} \int_{-T}^T \xi_T^2(t) dt = \frac{1}{4\pi T} \int_{-\infty}^{\infty} |\Xi_T(\omega)|^2 d\omega$$

La parte izquierda de la última expresión es proporcional a la potencia del proceso en el intervalo de análisis  $(-T, T)$ . Es más, cuando se analizan las señales aleatorias estacionarias, al hacer el intervalo  $T \rightarrow \infty$ , la potencia tiende al valor cuadrático medio dado en (3.11a), esto es,

$$\begin{aligned} E \left\{ \frac{1}{2T} \int_{-T}^T \xi_T^2(t) dt \right\} &= E \left\{ \frac{1}{4\pi T} \int_{-\infty}^{\infty} |\Xi_T(\omega)|^2 d\omega \right\} \\ \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T E \{ \xi_T^2(t) \} dt &= \lim_{T \rightarrow \infty} \frac{1}{4\pi T} \int_{-\infty}^{\infty} E \{ |\Xi_T(\omega)|^2 \} d\omega \\ \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T \overline{\xi_T^2(t)} dt &= \lim_{T \rightarrow \infty} \frac{1}{4\pi T} \int_{-\infty}^{\infty} E \{ |\Xi_T(\omega)|^2 \} d\omega \\ \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T \overline{\xi_T^2(t)} dt &= \frac{1}{2\pi} \int_{-\infty}^{\infty} \lim_{T \rightarrow \infty} \frac{E \{ |\Xi_T(\omega)|^2 \}}{2T} d\omega \\ E \{ \overline{\xi_T^2(t)} \} &= \frac{1}{2\pi} \int_{-\infty}^{\infty} \lim_{T \rightarrow \infty} \frac{E \{ |\Xi_T(\omega)|^2 \}}{2T} d\omega \end{aligned}$$

En el caso de los procesos ergódicos, el promedio de ensamble es igual al promedio de

tiempo, con lo que se obtiene,

$$\overline{\xi_T^2(t)} = \frac{1}{2\pi} \int_{-\infty}^{\infty} \lim_{T \rightarrow \infty} \frac{E \left\{ |\mathcal{E}_T(\omega)|^2 \right\}}{2T} d\omega \quad (3.22)$$

El operador dentro de la integral (3.22) es el promedio de tiempo del espectro de la observación de la señal aleatoria, denominado *densidad espectral de potencia* (DEP),

$$S_\xi(\omega) \triangleq \lim_{T \rightarrow \infty} \frac{E \left\{ |\mathcal{E}_T(\omega)|^2 \right\}}{2T} \quad (3.23)$$

Si la señal  $\xi(t)$  tiene unidades  $[V]$ , entonces la densidad  $S_\xi(\omega)$  tiene unidades  $[V^2/Hz]$ , que en correspondencia con (3.22), determina el valor cuadrático medio del proceso. Así,

$$\overline{\xi_T^2(t)} = \frac{1}{2\pi} \int_{-\infty}^{\infty} S_\xi(\omega) d\omega \quad (3.24)$$

La densidad espectral de potencia puede ser interpretada como la potencia media concentrada en los límites de una banda de frecuencia con banda de paso igual a  $1 Hz$  para una frecuencia central del espectro igual a  $\omega/2\pi [Hz]$ .

Atendiendo a las condiciones de realización físicas de los procesos aleatorios, la DEP tiene, entre otras, las siguientes propiedades:

- (a).  $S_\xi(\omega) \in \mathbb{R}$ , está determinada en el espacio de los reales,
- (b).  $0 \leq S_\xi(\omega) < \infty$ , es positiva semidefinida y acotada,
- (c).  $S_\xi(\omega) = S_\xi(-\omega)$ , es par. En consecuencia, la representación de  $S_\xi(\omega)$  en forma de funciones racionales, debe contener estrictamente polinomios de potencias pares:

$$S_\xi(\omega) = \frac{s_0(\omega^{2n} + a_{2n-2} + \dots + a_2\omega^2 + a_0)}{\omega^{2m} + b_{2m-2} + \dots + b_2\omega^2 + b_0}, \quad S_\xi(\omega) \in \mathbb{Q}$$

La expresión (3.24) se obtiene asumiendo el valor medio igual a 0 del proceso aleatorio. Sin embargo, se puede demostrar que la relación puede ser generalizada, obteniéndose:

$$\sigma_\xi^2 = \frac{1}{2\pi} \int_{-\infty}^{\infty} S_\xi(\omega) d\omega \quad (3.25)$$

**Transformada de Wiener-Jinchin.** La expresión (3.11c) define la función de correlación como la esperanza matemática del producto de dos funciones determinadas en el tiempo. Sin embargo, de la sección anterior, resulta que la DEP está relacionada con la esperanza matemática del producto de las TF de estas misma funciones. Entonces, debe existir una relación directa entre esas dos esperanzas matemáticas.

Sea una señal aleatoria con DEP, definida en (3.23),

$$S_{\xi}(\omega) = \lim_{T \rightarrow \infty} \frac{E \{ |\mathcal{E}_T(\omega)|^2 \}}{2T} = \lim_{T \rightarrow \infty} \frac{E \{ \mathcal{E}_T(\omega) \mathcal{E}_T(-\omega) \}}{2T}$$

Desarrollando las respectivas integrales de Fourier, se obtiene

$$\begin{aligned} S_{\xi}(\omega) &= \lim_{T \rightarrow \infty} \frac{1}{2T} E \left\{ \int_{-T}^T \xi_T(t_1) e^{j\omega t_1} dt_1 \int_{-T}^T \xi_T(t_2) e^{-j\omega t_2} dt_2 \right\} \\ &= \lim_{T \rightarrow \infty} \frac{1}{2T} E \left\{ \int_{-T}^T dt_2 \int_{-T}^T e^{-j\omega(t_2-t_1)} \xi_T(t_1) \xi_T(t_2) dt_1 \right\} \end{aligned}$$

El operador de promedio, atendiendo a la propiedad de linealidad de la TF, se puede incluir dentro de la integral:

$$S_{\xi}(\omega) = \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T dt_2 \int_{-T}^T e^{-j\omega(t_2-t_1)} E \{ \xi_T(t_1) \xi_T(t_2) \} dt_1$$

El operador de promedio de las funciones dentro de la integral corresponde a la función de correlación propia  $R_{\xi}(t_1, t_2)$  para la señal estacionaria en análisis, pero truncada  $\xi_T(t)$ . Realizando el cambio de notación,  $\tau = t_2 - t_1$ , por lo que  $dt_2 = d\tau$ , la anterior expresión toma la forma:

$$\begin{aligned} S_{\xi}(\omega) &= \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T-t_1}^{T-t_1} d\tau \int_{-T}^T e^{-j\omega\tau} R_{\xi}(t_1, t_1 + \tau) dt_1 \\ &= \int_{-\infty}^{\infty} \left( \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T R_{\xi}(t_1, t_1 + \tau) dt_1 \right) e^{-j\omega\tau} d\tau \end{aligned} \quad (3.26)$$

La expresión dentro del operador de promedio en la integral (3.26) corresponde a la función de correlación (3.11c), asumiendo la media del proceso  $\overline{\xi(t)} = 0$ . En otras palabras, la DEP de la señal aleatoria  $\xi(t)$  corresponde a la TF de la respectiva estimación en el tiempo de la función de correlación propia  $R_{\xi}(t, t + \tau)$ ,

$$S_{\xi}(\omega) = \mathcal{F} \{ \overline{R_{\xi}(t, t + \tau)} \}$$

Por cuanto en los procesos estacionarios se cumple que la función de correlación no depende del tiempo inicial del intervalo de análisis, esto es,  $\overline{R_{\xi}(t, t + \tau)} = R_{\xi}(\tau)$ , entonces, se obtiene que la DEP de un proceso estacionario en el sentido amplio es la TF de su

función de correlación:

$$S_{\xi}(\omega) = \mathcal{F} \{R_{\xi}(\tau)\} \quad (3.27)$$

De manera inversa se puede obtener que:

$$\mathcal{F}^{-1} \{S_{\xi}(\omega)\} = \mathcal{F}^{-1} \{\mathcal{F} \{R_{\xi}(\tau)\}\} = R_{\xi}(\tau) \quad (3.28)$$

El par conjugado de expresiones (3.27) y (3.28) conforma la *Transformada de Winner-Jinchin* que tiene un valor fundamental en el análisis de señales estacionarias, debido a que establecen la relación de su representación entre el dominio del tiempo (la función de correlación) y el dominio de la frecuencia (la DEP).

**Ejemplo 3.9.** Hallar la DEP para la función de correlación  $R_{\xi}(\tau) = \sigma_{\xi}^2 \exp(-\alpha|\tau|)$ .

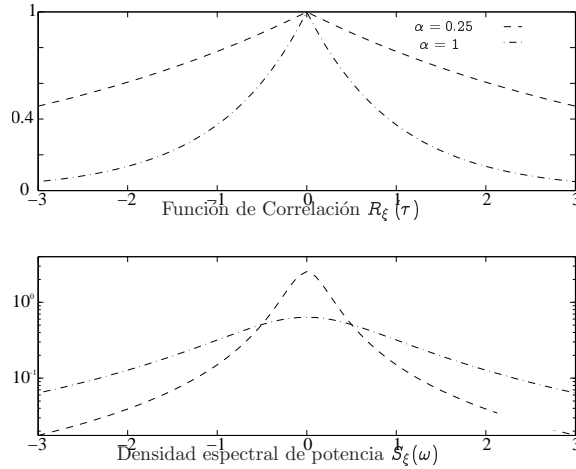
Empleando la relación (3.29a) se tiene:

$$S_{\xi}(\omega) = \frac{2}{\pi} \int_0^{\infty} \sigma_{\xi}^2 e^{-\alpha|\tau|} \cos \omega \tau d\tau = \frac{2\sigma_{\xi}^2}{\pi} \int_0^{\infty} e^{-\alpha\tau} d\tau \cos \omega \tau d\tau$$

La anterior integral se puede resolver por tabla:

$$S_{\xi}(\omega) = \frac{2\sigma_{\xi}^2}{\pi} \left( \frac{e^{-\alpha\tau}}{\alpha^2 + \omega^2} (\alpha \cos \omega \tau + \omega \sin \omega \tau) \right) \Big|_0^{\infty} = \frac{2\sigma_{\xi}^2}{\pi} \frac{\alpha}{\alpha^2 + \omega^2}$$

En la Figura 3.4 se presentan la función de correlación (parte superior) y la respectiva DEP (parte inferior); ambas calculadas para los casos de  $\alpha = 1$  y  $\alpha = 0.25$  y asumiendo un valor de la varianza unitario.



**Figura 3.4.** Resultados del ejemplo 3.9.

Empleando la propiedad de paridad de la función de correlación, las expresiones (3.27) y (3.28) se pueden llevar a las respectivas formas:

$$S_{\xi}(\omega) = 2 \int_0^{\infty} R_{\xi}(\tau) \cos \omega \tau d\tau \quad (3.29a)$$

$$R_{\xi}(\tau) = \frac{1}{\pi} \int_0^{\infty} S_{\xi}(\omega) \cos \omega \tau d\omega \quad (3.29b)$$

En la práctica, en calidad de valores de aleatoriedad para los procesos aleatorios, también se emplean los diferentes parámetros de su DEP, por ejemplo los siguientes:

– *Ancho de banda efectivo de espectro*

$$\Delta\omega_e = \frac{1}{S_{\xi}(0)} \int_{-\infty}^{\infty} S_{\xi}(\omega) d\omega$$

el cual está relacionado con el intervalo de correlación (3.15), por la expresión

$$\Delta\tau = \frac{1}{2} \frac{S_{\xi}(0)}{R_{\xi}(0)}$$

– Los siguientes momentos de frecuencia:

$$m_{1\omega} = \frac{2}{\sigma_{\xi}^2} \int_0^{\infty} \omega S_{\xi}(\omega) d\omega, \quad m_{2\omega} = \frac{2}{\sigma_{\xi}^2} \int_0^{\infty} \omega^2 S_{\xi}(\omega) d\omega, \quad \sigma_{\xi}^2 = \frac{2}{\sigma_{\xi}^2} \int_0^{\infty} (\omega - m_{1\omega})^2 S_{\xi}(\omega) d\omega$$

De otra parte, en el núcleo de la integral (3.26) está presente la función de correlación propia, que en el caso de considerar la función de correlación mutua (definida en (3.3b)), entonces, la función espectral respectiva es definida como la *densidad espectral de potencia mutua*:

$$S_{\xi\eta}(\omega) = \int_{-\infty}^{\infty} R_{\xi\eta}(\tau) e^{-j\omega\tau} d\tau \quad (3.30)$$

Cabe anotar, que cuando las señales aleatorias  $\xi(t)$  y  $\eta(t)$  tienen correlación nula, esto es,  $R_{\xi\eta}(\tau) = 0$ , la correspondiente DEP mutua también es de contenido 0.

La relación entre las densidades de potencia  $S_{\xi\eta}(\omega)$  y  $S_{\eta\xi}(\omega)$  se deduce a partir del análisis de las partes real e imaginaria de la definición (3.30),

$$S_{\xi\eta}(\omega) = \int_{-\infty}^{\infty} R_{\xi\eta}(\tau) \cos \omega \tau d\tau - j \int_{-\infty}^{\infty} R_{\xi\eta}(\tau) \sin \omega \tau d\tau, \quad (3.31)$$

donde las respectivas funciones de correlación mutua se definen como

$$R_{\xi\eta}(\tau) = \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T \xi(t) \eta(t + \tau) dt, \quad (3.32a)$$

$$R_{\eta\xi}(\tau) = \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T \eta(t) \xi(t + \tau) dt \quad (3.32b)$$

En la última integral, (3.32b), se realizan los siguientes cambios de variables:  $\lambda = t + \tau$ , luego,  $t = \lambda - \tau$ ,  $dt = d\lambda$ , con lo cual,

$$R_{\eta\xi}(\tau) = \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T \eta(\lambda - \tau) \xi(\lambda) d\lambda$$

regresando de nuevo a la variable  $t$ , se obtiene,

$$R_{\eta\xi}(\tau) = \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T \eta(t) \xi(t - \tau) dt$$

Al comparar la última expresión con la definición (3.32b) se observa el cumplimiento de la igualdad  $R_{\xi\eta}(\tau) = R_{\eta\xi}(-\tau)$ , que al sustituir en la (3.31), resulta en

$$S_{\eta\xi}(\omega) = \int_{-\infty}^{\infty} R_{\xi\eta}(-\tau) \cos \omega\tau d\tau - j \int_{-\infty}^{\infty} R_{\xi\eta}(-\tau) \sin \omega\tau d\tau$$

Sea  $-\tau = \lambda$ , entonces

$$\begin{aligned} S_{\eta\xi}(\omega) &= \int_{-\infty}^{\infty} R_{\xi\eta}(\lambda) \cos \omega\lambda d(-\lambda) - j \int_{-\infty}^{\infty} R_{\xi\eta}(\lambda) \sin(-\omega\lambda) d(-\lambda) \\ &= \int_{-\infty}^{\infty} R_{\xi\eta}(\lambda) \cos \omega\lambda d\lambda + j \int_{-\infty}^{\infty} R_{\xi\eta}(\lambda) \sin \omega\lambda d\lambda \end{aligned}$$

de lo cual, finalmente, se obtiene que

$$S_{\eta\xi}(\omega) = S_{\xi\eta}^*(\omega)$$

**Ruido blanco Gaussiano (RBG).** Se denomina *ruido blanco Gaussiano* al modelo de señal aleatoria que se define convencionalmente como una señal ergódica con FDP Gaussiana y DEP constante en todo el dominio de la frecuencia (Figura 3.5(b)),

$$S(\omega) = N_0, \omega \in (-\infty, \infty). \quad (3.33)$$

El modelo (3.33), denominado *ruido blanco* por su analogía espectral con la luz blanca, tiene función de correlación propia obtenida mediante la TF (Figura 3.5(a)):

$$R(\tau) = \int_{-\infty}^{\infty} \frac{N_0}{2} e^{j2\pi f\tau} df = \frac{N_0}{2} \delta(\tau) \quad (3.34)$$

Del modelo (3.34), se observa que  $R(\tau) = 0, \forall \tau \neq 0$ , de tal manera que cualquier par de diferentes valores tomados del RBG no son correlacionadas y, por lo tanto, estadísticamente son independientes.

Cabe anotar que el modelo (3.34) implica que la varianza del RBG  $\sigma^2 \rightarrow \infty$ .

En realidad, todos los dispositivos de procesos de señales eléctricas poseen un ancho de banda finito  $\Delta\omega$ , y en consecuencia cualquier clase de ruido también tendrá un ancho de banda finito a la salida de todo dispositivo. Cuando se asume la constancia de la DEP en el ancho de banda finito  $\Delta\omega < \infty$ , se tiene el siguiente modelo de ruido:

$$\begin{cases} S(\omega) = N_0, (-\Delta\omega < \omega < \Delta\omega), \\ R(\tau) = N_0 \Delta\omega \text{sinc}(2\Delta\omega\tau), \end{cases} \quad (3.35)$$

Puesto que el ancho de banda espectral resultante (3.35) es menor que del ruido blanco, al modelo filtrado se le conoce como *ruido rosado* o *blanco de banda finita*, para el cual además se cumple que su potencia de salida es finita e igual a  $N_0 \Delta\omega$  y su valores instantáneos están correlacionados.

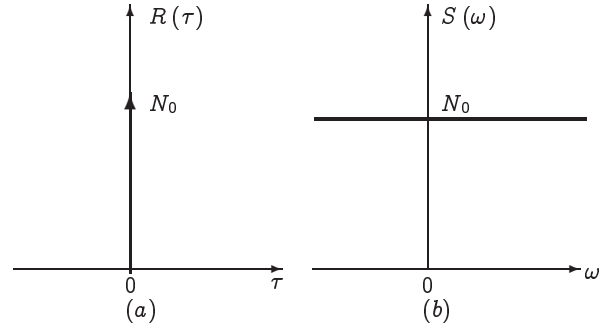
En la práctica, se define el *ancho de banda equivalente del ruido*  $\Delta\omega_e$ :

$$\Delta\omega_e = \frac{1}{H_0} \int_0^{\infty} |H(f)|^2 df, \quad (3.36)$$

donde  $H_0 = |H(f)|_{\max}$  es la ganancia de voltaje del filtro en la frecuencia central. Luego, la potencia de ruido promedio filtrado es:

$$N = \int_{-\infty}^{\infty} |H(f)|^2 N_0 df = N_0 \int_0^{\infty} |H(f)|^2 df = N_0 H_0 \Delta\omega_e,$$

La última ecuación muestra el efecto del filtro separado en dos partes: la selectividad de frecuencia relativa, representada por medio de  $\Delta\omega_e$ , y la ganancia de potencia representada por medio de  $H_0$ . Por definición, el ancho de banda equivalente de ruido de un filtro ideal es su ancho de banda real. En los filtros reales,  $\Delta\omega_e$  es algo mayor que el ancho de banda a 3 dB [14].



**Figura 3.5.** Modelo del ruido blanco Gaussiano

### 3.1.5. Convergencia y continuidad de procesos aleatorios

El análisis de modelos dinámicos, excitados por procesos aleatorios, se basa bien en la solución de ecuaciones diferenciales que describen los sistemas, bien en el análisis de otras características equivalentes. En cualquier caso, como ocurre en el estudio de excitaciones determinísticas, es necesario determinar el alcance de las definiciones de los conceptos de la continuidad, la diferenciabilidad e integrabilidad de los procesos aleatorios.

**Convergencia.** Sea la sucesión de valores aleatorios  $\{\xi_k : k = 1, \dots, n\}$ , para la cual, la convergencia se puede determinar en alguno de los siguientes sentidos [15]:

1. La sucesión  $\{\xi_k\}$  converge al valor aleatorio *con probabilidad de 1* (casi seguramente) si se cumple que,

$$P\{\xi_n \rightarrow \xi\}, \text{ para } n \rightarrow \infty$$

2. La sucesión  $\{\xi_k\}$  converge a  $\xi$  *en el sentido probabilístico*, si para cualquier  $\varepsilon > 0$ ,

$$P\{\xi_k - \xi > \varepsilon\} = 0, \text{ para } n \rightarrow \infty$$

3. La sucesión de valores aleatorios  $\{\xi_k\}$  converge a  $\xi$  en el *sentido cuadrático medio*, cuando,

$$\lim_{n \rightarrow \infty} E\{|\xi_n - \xi|^2\} = 0$$

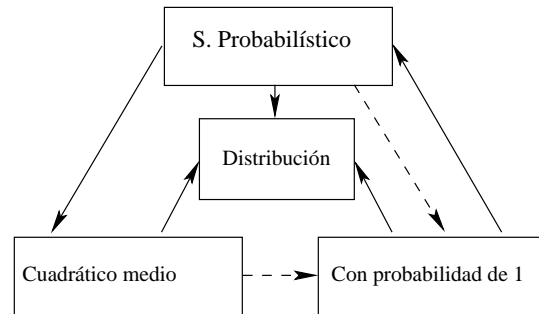
que frecuentemente se nota como,  $\text{l.i.m.}_{n \rightarrow \infty} \xi_n = \xi$ .

4. Se dice que  $\xi_n$  converge a  $\xi$  *por distribución*, si dadas las respectivas funciones de distribución,  $F_n(x)$  y  $F(x)$ ,  $\forall x$ , se cumple que  $F_n(x) \rightarrow F(x)$ , asumiendo  $n \rightarrow \infty$ .

Se considera que la convergencia con probabilidad de 1 implica la convergencia en el sentido probabilístico, mientras, la convergencia por valor cuadrático medio también implica la convergencia en el sentido probabilístico, como se muestra en la Figura 3.6, en la cual con líneas a trozos se representa la relación restringida a ciertas clases de sucesiones aleatorias.

En la práctica, el valor límite  $\xi$  no se conoce, entonces, similar al caso de sucesiones determinísticas como criterio de convergencia se emplea el *criterio de Cauchy*,

$$\lim_{n \rightarrow \infty} |\xi_{n+m} - \xi_n| \rightarrow 0, \forall m$$



**Figura 3.6.** Relación entre las definiciones de convergencia probabilística



Cuando el anterior límite existe, entonces se da la convergencia para cualquiera de sus formas anteriormente definidas. Las definiciones dadas de convergencia para sucesiones aleatorias se pueden extender a los procesos continuos aleatorios  $\{\xi(t), t \in T\}$ , teniendo en cuenta que las definiciones no se restringen solamente a una observación en particular del proceso, sino al conjunto de trayectorias en su totalidad.

**Continuidad.** Mediante las definiciones anteriores de convergencia, también es posible determinar la continuidad, así por ejemplo, una señal aleatoria  $\xi(t)$  se define como continua en  $t$  con probabilidad 1, cuando se cumple que:

$$\lim_{\tau \rightarrow 0} \xi(t + \tau) = \xi(t)$$

Sin embargo, no debe entenderse que la continuidad definida para los procesos aleatorios asegura la continuidad de cualquiera de sus trayectorias. Por ejemplo, un proceso de Poisson con FDP, dada por la expresión (3) del ejemplo 2.8, para un valor dado de  $t = t_0$ , es un proceso continuo, pero cualquiera de sus observaciones no lo es.

Con frecuencia, el criterio más aceptado de convergencia para un proceso aleatorio es el del valor cuadrático medio, el cual se asume que es un *proceso de segundo orden*, cuando se cumple que  $E\{\xi^2(t)\} < \infty, \forall t \in T$ .

Un proceso aleatorio de segundo orden,  $\xi(t), t \in T$ , es continuo en  $t$  en el sentido cuadrático medio, si se cumple que,

$$\lim_{\tau \rightarrow 0} E\{(\xi(t + \tau) - \xi(t))^2\} = 0 \quad (3.37)$$

Además, de (3.37) se obtiene que

$$\begin{aligned} E\{\xi(t + t_0), \xi(t)\} &= E\{(\xi(t + t_0) - m_{1\xi}(t) - \xi(t) - m_{1\xi}(t))^2 + \\ &\quad + 2(\xi(t + t_0) - \xi(t))((m_{1\xi}(t + t_0) - m_{1\xi}(t)) - (m_{1\xi}(t + t_0) - m_{1\xi}(t)))^2\} \\ &= R_\xi(t + t_0, t + t_0) - 2R(t + t_0, t + t_0) - 2R(t + t_0, t) + R(t, t) + \\ &\quad + (m_{1\xi}(t + t_0) - m_{1\xi}(t))^2 \end{aligned}$$

Por lo tanto, para la continuidad del proceso  $\xi(t)$  en  $t$ , es necesario que tanto su función de correlación  $R_\xi(t, t)$  como su valor medio  $m_\xi(t)$  sean funciones continuas. De otra parte, teniendo en cuenta que el valor cuadrático medio es positivo definido, entonces, la condición suficiente de continuidad corresponde a,

$$\begin{aligned} &E\{(\xi(t + t_0) - m_{1\xi}(t + t_0) - \xi(t) - m_{1\xi}(t))^2\} \\ &= R(t + t_0, t + t_0) - 2R(t + t_0, t) + R(t, t) \geq 0 \end{aligned}$$

**Diferenciabilidad.** Un proceso aleatorio  $\xi(t), t \in T$ , de segundo orden se define como diferenciable en el sentido cuadrático medio en el punto  $t_0 \in T$ , si se cumple que para el

límite

$$\lim_{\tau \rightarrow 0} \frac{\xi(t_0 + \tau) - \xi(t_0)}{\tau} = \xi'(t_0)$$

existe un valor  $\lim_{\tau \rightarrow 0} E \left\{ \left( \frac{\xi(t_0 + \tau) - \xi(t_0)}{\tau} - \xi'(t_0) \right)^2 \right\} = 0$ . Cuando esta condición se cumple para todo  $t \in T$ , entonces se dice que el proceso aleatorio es *diferenciable*.

La condición necesaria de diferenciabilidad del proceso en el punto  $t_0 \in T$  se halla analizando la siguiente sucesión de valores contiguos:

$$\begin{aligned} & E \left\{ \left( \frac{\xi(t_0 + \tau_1) - \xi(t_0)}{\tau_1} - \frac{\xi(t_0 + \tau_2) - \xi(t_0)}{\tau_2} \right)^2 \right\} \\ &= E \left\{ \frac{\xi(t_0 + \tau_1) - \xi(t_0)}{\tau_1} \frac{\xi(t_0 + \tau_1) - \xi(t_0)}{\tau_1} - 2 \frac{\xi(t_0 + \tau_1) - \xi(t_0)}{\tau_1} \frac{\xi(t_0 + \tau_2) - \xi(t_0)}{\tau_2} + \right. \\ &\quad \left. + \frac{\xi(t_0 + \tau_2) - \xi(t_0)}{\tau_2} \frac{\xi(t_0 + \tau_2) - \xi(t_0)}{\tau_2} \right\} = \\ &= \frac{R_\xi(t_0 + \tau_1, t_0 + \tau_2) - R(t_0 + \tau_1, t_0) - R(t_0, t_0 + \tau_2) + R_\xi(t_0, t_0)}{\tau_1 \tau_2} + \\ &\quad + \frac{m(t_0 + \tau_1) - m(t_0)}{\tau_1} \frac{m(t_0 + \tau_2) - m(t_0)}{\tau_2} \end{aligned} \quad (3.38)$$

Debido a la suposición sobre la diferenciabilidad del valor medio, la cual implica la existencia de la segunda derivada de la función  $R_\xi(t_1, t_2)$ , en los puntos de tiempo  $t_0$  y  $t_1 = t_2 - t_0$ , esto es,

$$\lim_{\tau_1, \tau_2 \rightarrow 0} E \left\{ \frac{\xi(t_0 + \tau_1) - \xi(t_0)}{\tau_1} \frac{\xi(t_0 + \tau_2) - \xi(t_0)}{\tau_2} \right\} = \frac{\partial^2 R(t_1, t_2)}{\partial t_1 \partial t_2} \Big|_{t_1=t_2=t_0} + m'(t_0)m'(t_0)$$

entonces, la parte derecha de la igualdad (3.38) es finita, para  $\tau_1, \tau_2 \rightarrow 0$ , se cumple que

$$E \left\{ \left( \frac{\xi(t_0 + \tau_1) - \xi(t_0)}{\tau_1} - \frac{\xi(t_0 + \tau_1) - \xi(t_0)}{\tau_1} \right)^2 \right\} \rightarrow 0$$

La condición suficiente de diferenciabilidad del proceso se halla asumiendo  $\tau_1 = \tau_2$ , con lo cual, la ec. (3.38) es la suma de dos términos positivos. Si la parte derecha de la igualdad tiende a 0, es suficiente que cada uno de sus términos de la derecha tiendan también a 0.

**Integrabilidad.** Sean un proceso aleatorio  $\xi(t)$  y la función determinística  $f(t)$ , tales que existan en el intervalo  $[a, b]$ , sobre el que se define la malla de puntos en el tiempo de la forma:  $a = t_0 < t_1 < \dots < t_n = b$ , sobre la cual se analiza la suma:

$$\mathcal{J}(n) = \sum_{i=1}^n f(t_i) \xi(t_i) (t_i - t_{i-1})$$

Si para un  $\max_{1 \leq i \leq n} (t_i - t_{i-1})$ , la suma converge en algún vector límite, que a propósito es otra variable aleatoria, entonces ese límite se denomina *integral del proceso aleatorio*  $\xi(t)$

$$\mathcal{J} = \int_a^b f(t)\xi(t)dt$$

La convergencia de las sumas  $\mathcal{J}(n)$  se analiza en el sentido cuadrático medio, esto es,  $E\{(\mathcal{J}(n) - \mathcal{J})^2\} \rightarrow 0$ ,  $n \rightarrow \infty$ , mientras el límite  $\mathcal{J}$  es la integral cuadrática media. Un proceso aleatorio de segundo orden,  $\xi(t)$ ,  $t \in [a, b]$  con  $m_{1\xi}(t)$  y  $R_\xi(t_1, t_2)$  es integrable por Riemann, cuando existen las integrales,

$$\int_a^b f(t)m_{1\xi}(t)dt, \int_a^b \int_a^b f(t_1)f(t_2)R_\xi(t_1, t_2)dt_1dt_2$$

en este caso, se tiene que

$$E \left\{ \int_a^b f(t)\xi(t)dt \right\} = \int_a^b f(t)m_{1\xi}(t)dt$$

$$E \left\{ \int_a^b \int_a^b f(t_1)f(t_2)\xi(t_1)\xi(t_2)dt_1dt_2 \right\} = \int_a^b \int_a^b f(t_1)f(t_2)R(t_1, t_2)dt_1dt_2 + \left( \int_a^b f(t)m_{1\xi}(t)dt \right)^2$$

Algunas veces, es necesario analizar la integral estocástica de Stieltjes,  $\int_a^b f(t)d\xi(t)$  que se define por el límite cuando  $\max_{1 \leq i \leq n} (t_i - t_{i-1}) \rightarrow 0$ , por la siguiente suma:

$$\sum_{i=1}^n f(t_i) (\xi(t_i) - \xi(t_{i-1}))$$

Es de anotar que el concepto de continuidad y diferenciabilidad de un proceso aleatorio no son equivalentes. La exigencia sobre la diferenciabilidad de un proceso aleatorio es más fuerte y restrictiva, que la continuidad del mismo. Por ejemplo, el proceso estacionario con función de correlación  $R_\xi(\tau) = \sigma_\xi^2(-\alpha|\tau|)$  es continuo, pero en general se demuestra que no es diferenciable.

### Problemas

**Problema 3.1.** Hallar el primer momento inicial  $m_{1\xi}(t)$  y la varianza  $\sigma_\xi^2(t)$  del proceso con FDP,

$$p(\xi, t) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{\xi^2}{2} e^{2\alpha t} + \alpha t\right)$$

**Problema 3.2.** Hallar la función de correlación y determinar la condición de estacionariedad de la señal aleatoria  $\eta(t) = y(t)\xi(t)$ , siendo  $y(t)$  una función no aleatoria.

**Problema 3.3.** Calcular los intervalos de correlación para las siguientes funciones de correlación  $R_\xi(\tau)$ : 1.  $a \exp(-\alpha|\tau|)$  2.  $a \exp(-\alpha^2\tau^2)$  3.  $a \exp(-\alpha^2\tau^2) \cos \omega_0\tau$  4.  $a \operatorname{sinc} \alpha\tau$

**Problema 3.4.** Hallar la función de correlación propia de la señal aleatoria periódica,

$$\xi(t) = \sum_{n=1}^{\infty} \frac{1}{n} \sin n\Omega t$$

**Problema 3.5.** Hallar la función de correlación  $R_\xi(\tau)$  de un proceso aleatorio estacionario con DEP dada así:

$$S_\xi(\omega) = \begin{cases} N/2, & |\omega_1| \leq \omega \leq |\omega_2| \\ 0, & \text{otros valores de } \omega \end{cases}$$

**Problema 3.6.** Hallar la DEP  $S_\xi(\omega)$  del proceso aleatorio  $\xi$ , cuya función de correlación es igual a

$$R_\xi(\tau) = \begin{cases} \sigma_\xi^2 \left(1 - \frac{\tau}{T}\right), & |\tau| \leq T \\ 0, & \text{otros valores de } t \end{cases}$$

**Problema 3.7.** Demostrar que para un proceso estacionario dado, el cambio de escala  $a$  en el argumento de la función de correlación, corresponde al siguiente cambio de escala en la DEP,

$$R_\xi(a\tau) \Leftrightarrow \frac{1}{a} S_\xi\left(\frac{\omega}{a}\right)$$

**Problema 3.8.** Demostrar que, en forma general, la DEP conjunta  $S_{\xi\eta}(\omega)$  no es una función par.

### 3.2. Análisis experimental de señales estacionarias

Sea una señal aleatoria y variable en el tiempo  $\xi(t)$ , entonces la mayoría de sus valores de aleatoriedad (momentos iniciales, centralizados, función de correlación, etc.), pueden ser también función del tiempo. Excepto, la densidad espectral de potencia que varía en términos de la frecuencia angular  $\omega = 2\pi f$ . El análisis experimental de la variable aleatoria  $\xi(t)$  para la estimación de alguna de sus valores de aleatoriedad  $\tilde{\theta}(t)$ , exige la realización de  $n$  experimentos y la obtención de una serie de observaciones  $\xi_1(t), \xi_2(t), \dots, \xi_n(t)$ . Como resultado se obtiene la estimación en la forma:

$$\tilde{\theta}(t) = g_{\theta}(\xi_1(t), \dots, \xi_n(t)) \quad (3.39)$$

**Tareas básicas de estimación.** En el análisis estadístico, de acuerdo a la información a priori dispuesta, se aceptan tres formas básicas de estructura de una señal  $\xi(t)$ ,  $t \in T$ , considerada estacionaria en el sentido amplio:

1.  $\xi(t) = \xi_0(t)$ ,  $t \in T$
2.  $\xi(t) = m_{\xi} + \xi_0(t)$ ,  $t \in T$
3.  $\xi(t) = \sum_{k=1}^r \theta_k a_k(t) + \xi_0(t)$ ,

donde  $\xi_0(t)$  es un proceso estacionario en el sentido amplio con media igual a cero.

Sea  $\xi_k(t)$ ,  $t \in T_a$  una trayectoria de la variable aleatoria  $\xi(t)$ , observada en el intervalo de análisis de tiempo  $T_a$ , donde  $T_a$  puede ser un segmento de tiempo  $T_a \in [a, b]$  para el caso de variables continuas, o una sucesión de valores definidos en un tiempo de observación  $T_a \in \{t_k, k = 1, \dots, n\}$ , para el caso de una variable aleatoria discreta.

En concordancia con lo anterior, se establecen las siguientes tareas básicas de estimación en los procesos estacionarios:

1. Estimación de la función de DEP,  $S_{\xi}(\omega)$ , para proceso  $\xi(t)$ ,
2. Conocida la DEP  $S_{\xi}(\omega)$  del proceso  $\xi_0(t)$ , se exige estimar su valor medio  $m_{\xi}$ ,
3. Conocida la DEP del proceso  $\xi_0(t)$ , se exige estimar los parámetros  $\theta_1, \dots, \theta_r$  de la regresión  $\sum_{k=1}^r \theta_k a_k(t)$ , cuando la función  $a_k(t)$  se asume conocida.

Sea  $\tilde{\theta} = g_{\theta}(\xi(t))$ ,  $t \in T_a$ , la estadística destinada a la solución de cualquiera de los tres casos básicos anteriores. De todas las posibles estadísticas  $\tilde{\theta}$  se escogen aquellas que presenten las propiedades expuestas en el numeral §2.2, en particular, las siguientes:

- (a). *Linealidad.* El funcional  $g_{\theta}(\cdot)$  debe ser lineal,
- (b). *Ausencia de sesgo.* Si la estimación del parámetro  $\theta$  se realiza mediante la estadística  $\tilde{\theta}$ , se exige que  $E\{\tilde{\theta}\} = \theta$ ,

- (c). *Consistencia.* La estadística  $\tilde{\theta}$  debe converger en el sentido probabilístico al valor  $\theta$ , en la medida en que se aumente el intervalo de observación,
- (d). *Efectividad.* La estadística  $\tilde{\theta}$  debe tener la menor de todas las varianzas posibles dentro de las estadísticas de una clase dada.

### 3.2.1. Estimación de momentos

**Estimación del valor medio.** En concordancia con la clase de estimadores lineales y sin sesgo, propuestos en (3.8) para los procesos ergódicos, se define el siguiente estimador:

$$\tilde{m}_{1\xi} = \overline{\xi(t)} = \int_a^b w_m(t) \xi(t) dt$$

donde la función  $w_m(t)$  es la función ventana, que cumple la condición (3.9).

La ausencia de sesgo del anterior estimador se comprueba de lo siguiente:

$$\begin{aligned} E \left\{ \overline{\xi(t)} \right\} &= E \left\{ \int_a^b w_m(t) \xi(t) dt \right\} = \int_a^b w_m(t) E \{ \xi(t) \} dt = \overline{\xi(t)} \int_a^b w_m(t) dt \\ &= \overline{\xi(t)} \end{aligned}$$

La precisión en la aproximación de la estimación,  $\tilde{m}_{1\xi} \approx \overline{\xi(t)}$ , se caracteriza por la respectiva varianza de la potencia de error descrita en (2.63) [23]:

$$\begin{aligned} E \left\{ \left( m_{1\xi} - E \left\{ \overline{\xi(t)} \right\} \right)^2 \right\} &= E \left\{ m_{1\xi}^2 \right\} - E \left\{ \overline{\xi^2(t)} \right\} \\ &= 2 \int_0^T R_\xi(\tau) \int_0^{T-\tau} w_m(t) w_m(t+\tau) dt d\tau \end{aligned} \quad (3.40)$$

La precisión de la estimación depende básicamente de la función ventana, cuya optimización por el criterio del mínimo error cuadrático medio conlleva al aumento significativo del costo computacional. Por esto, en la práctica se emplea la ventana más simple,

$$w_m(t) = \begin{cases} \frac{1}{T}, & 0 \leq t \leq T \\ 0, & 0 > t > T \end{cases} \quad (3.41)$$

con lo que la estimación de la media, para señales aleatorias continuas, es de la forma:

$$\overline{\xi(t)} = \int_0^T \xi(t) dt \quad (3.42)$$

La estimación del valor medio para una señal aleatoria discreta  $\{\xi[k] : k = 1, \dots, N\}$ , de

forma similar, se realiza discretizando la expresión (3.42),

$$\tilde{m}_{1\xi} = \sum_{k=1}^N \xi[k] w_m[k], \quad (3.43)$$

donde la función de peso discreta cumple la condición,  $\sum_{k=1}^N w_m[k] = 1$ . En los argumentos de la serie (3.43) se asume la normalización de las bases de tiempo,  $x[kT_d] = x[k]$ .

En la práctica, es frecuente el uso de la función ventana rectangular,  $w_m[k] = 1/N$ ,  $k = 1, \dots, N$ , para la cual, la varianza del error (3.40) toma la forma:

$$\sigma_\varepsilon^2 = \frac{\sigma_\xi^2}{N^2} \left( N + 2 \sum_{k=1}^{N-1} (N - k) \rho_\xi[k] \right)$$

donde  $R_\xi(\tau) = \sigma_\xi^2 \rho_\xi(\tau)$ . La relación  $\sigma_\varepsilon^2 / \sigma_\xi^2$  debe corresponder al mínimo valor posible, que es inversamente proporcional al tamaño  $N$  de la sucesión  $\xi[k]$ . Por lo tanto, el valor  $N$  corresponde al compromiso entre el costo computacional y la precisión del estimador.

**Estimación de la varianza.** Cuando el valor medio  $\overline{\xi(t)}$  del proceso aleatorio es 0, la estimación de la varianza coincide con el valor cuadrático medio y puede realizarse acorde con la definición (3.11a), teniendo en cuenta la clase de estimadores propuestos en (3.8),

$$\tilde{m}_{2\xi} = \overline{\xi^2(t)} = \int_0^T w_m(t) \xi^2(t) dt$$

Si el valor medio no es conocido, entonces la estimación de la varianza será

$$\tilde{\sigma}_\xi^2 = \int_0^T w_\sigma(t) \left( \xi^2(t) - \overline{\xi(t)} \right)^2 dt = \int_0^T w_\sigma(t) \left( \xi^2(t) - \int_0^T w_m(t) \xi(t) dt \right)^2 dt \quad (3.44)$$

Si se asume la ventana  $w_m$  de (3.41), además al exigir la constancia de la otra ventana,  $w_\sigma = \text{const}$ , para asegurar la ausencia de sesgo en (3.44), la función  $w_\sigma$  toma la forma

$$w_\sigma(t) = \frac{1}{T(1 - k_\sigma)} \quad (3.45)$$

El valor de la constante  $k_\sigma$  se determina experimentalmente [23].

**Estimación de la función de correlación.** Se tienen diferentes métodos de estimación de la función de correlación, uno de los más empleados es el siguiente:

$$\tilde{R}_\xi(\tau) = \int_0^{T-\tau} w_R(t, \tau) (\xi(t) - \tilde{m}_{1\xi}) (\xi(t + \tau) - \tilde{m}_{1\xi}) dt \quad (3.46)$$

donde  $w_R(t, \tau)$  es la ventana que se propone para la estimación de función de correlación.

Cuando se tiene que  $\bar{\xi}(t) = 0$ , la estimación (3.46) no presenta sesgo:

$$\begin{aligned} E \{ \tilde{R}_\xi(\tau) \} &= E \left\{ \int_0^{T-\tau} w_R(t, \tau) \xi_0(t) \xi_0(t + \tau) dt \right\} = \int_0^T w_R(t, \tau) E \{ \xi_0(t) \xi(t + \tau) \} dt \\ &= R_\xi(\tau) \end{aligned}$$

Además, la varianza de la estimación resulta ser igual a

$$\begin{aligned} \sigma_R^2 &= \int_0^{T-\tau} \int_0^{T-\tau} w_R(\lambda_1, \tau) w_R(\lambda_2, \tau) \times \\ &\quad \times \left( R_\xi^2(\lambda_1 - \lambda_2) + R_\xi(\lambda_1 - \lambda_2 - \tau) R_\xi(\lambda_1 - \lambda_2 + \tau) \right) d\lambda_1 d\lambda_2 \\ &= \int_{-(T-\tau)}^{T-\tau} w_R^2(\lambda, \tau) \left( R_\xi^2(\lambda) + R_\xi^2(\lambda - \tau) R_\xi^2(\lambda + \tau) \right) d\lambda \end{aligned}$$

De la anterior expresión se deduce que para determinar la varianza de la estimación para la función de correlación hay que conocer la misma señal aleatoria. Por esta razón, la determinación de la precisión de la estimación se puede realizar después de realizar el procesamiento, además, hay que tener en cuenta el hecho concreto de que la misma varianza es un proceso aleatorio. El desconocimiento a priori de la función de correlación, hace que en la mayoría de los casos no se plantee el problema de optimización de la función ventana de estimación  $w_R$ . Aunque, en el procedimiento de prueba de hipótesis, cuando se comprueba la pertenencia o no de una función de correlación a una clase dada, puede ser necesario la determinación de la ventana óptima de estimación.

En la práctica, la estimación de la función de correlación se realiza a partir de la ventana similar a la propuesta en (3.45):

$$\tilde{R}_\xi(\tau) = \frac{1}{(T-\tau)} \int_0^{T-\tau} (\xi(t) - m_{1\xi}) (\xi(t + \tau) - m_{1\xi}) dt \quad (3.47)$$

para la cual, se tiene el siguiente valor medio

$$\begin{aligned} E \{ \tilde{R}_\xi(\tau) \} &= R_\xi(\tau) - \frac{2}{T(T-\tau)} \int_0^T \left( 1 - \frac{\lambda}{T} \right) (\tau R_\xi(\tau) + T R_\xi(\lambda - \tau)) d\lambda + \\ &\quad + \frac{1}{T(T-\tau)} \int_0^T (T + \tau - 2\lambda) (R_\xi(\tau) + R_\xi(\lambda - \tau)) d\tau \end{aligned}$$

que muestra un sesgo en la estimación  $\tilde{R}_\xi$ , el cual se disminuye mediante el factor de corrección:  $1/(1 - f(R_\xi(\tau)))$ , donde la dependencia  $f(\cdot)$  se establece empíricamente.



De otra parte, la estimación del índice de correlación  $\rho(\tau)$ , (2.46), se realiza mediante la relación de los valores,  $\tilde{\rho}_\xi(\tau) = \tilde{R}_\xi(\tau) / \tilde{R}_\xi(0)$ , que resulta ser asintóticamente no sesgada:

$$E \{ \tilde{\rho}_\xi(\tau) - \rho_\xi(\tau) \} \approx \frac{2}{T} \int_{-\infty}^{\infty} \left( \rho^2(t) \rho(\tau) - \rho(t) \rho(t-\tau) \right) dt$$

En realidad, la estimación del índice de correlación de un proceso ergódico, inclusive siendo conocido el valor medio del proceso, presenta un sesgo, cuyo valor disminuye en la medida en que aumenta la longitud de la observación de la señal aleatoria en análisis.

Finalmente, en el caso de tener sucesiones estacionarias aleatorias, la estimación de la función de correlación se toma de la siguiente forma:

$$\tilde{R}_\xi[k] = \frac{1}{N-k} \sum_{l=1}^{N-k} \xi_0[l] \xi_0[l-k] \quad (3.48)$$

### 3.2.2. Estimación de los coeficientes en la descomposición K-L

En la sección §3.1.3 se plantea la búsqueda de un conjunto óptimo de funciones base con dimensión  $p$ , dado en  $L^2(T)$ , para la representación discreta finita de la forma (1.4) a partir de diversas observaciones de un proceso aleatorio  $\xi(t)$ ,  $t \in T$ , de tal manera que la norma del error en  $L^2(T)$ , promediada sobre el conjunto de observaciones, sea la menor posible.

En particular, de (3.18) se tiene que:

$$\begin{aligned} \bar{\varepsilon}_T^2 &= \int_T E \{ \xi(t) \xi^*(t) \} dt - \sum_{i=1}^p \int_T \int_T E \{ \xi(t) \xi^*(t) \} \phi_i^*(t) \phi_i(\tau) dt d\tau \\ &= \int_T R_\xi(t, \tau) dt - \sum_{i=1}^p \int_T \int_T R_\xi(t, \tau) \phi_i^*(t) \phi_i(\tau) dt d\tau \end{aligned} \quad (3.49)$$

El primer término de (3.49) no depende del conjunto base  $\phi = \{\phi_i\}$ , por lo que la tarea se resume a hallar las  $p$  funciones ortogonales que maximizan el segundo término:

$$\sum_{i=1}^p \int_T \int_T R_\xi(t, \tau) \phi_i^*(t) \phi_i(\tau) d\tau = \sum_{i=1}^p \langle \mathcal{A}_\xi \{ \phi_i(t) \}, \phi_i(t) \rangle$$

que corresponde a una suma de funcionales cuadráticos. La condición de valor máximo se encuentra, al tener en cuenta que el núcleo del operador integral  $\mathcal{A}_\xi$ , definido como,

$$\mathcal{A}_\xi \{ \phi(t) \} = \int_T R_\xi(t, \tau) \phi(\tau) d\tau \quad (3.50)$$

que corresponde a la función de correlación propia de un proceso con valor medio cuadrático finito. El operador (3.50) tiene las siguientes propiedades:

- (a). Integración cuadrática del núcleo,  $\int_T \int_T |\mathcal{A}(t, \tau)|^2 dt d\tau < \infty$ , en particular, en forma de un operador del tipo Hilbert-Schmidt.
- (b). Simetría,  $\mathcal{A}(t, \tau) = \mathcal{A}^*(t, \tau)$ , que resulta de las propiedades de la función de correlación propia dada en la sección §3.1.2,  $R_\xi(t, \tau) = R_\xi^*(\tau, t)$ .
- (c). Naturaleza positiva semidefinida,  $\langle \mathcal{A}\{\phi\}, \phi \rangle = E \{ \|\xi, \phi\|^2 \} \geq 0$ .

De las anteriores propiedades resulta lo siguiente:

1. Los valores propios de la solución (3.50) conforman una sucesión cuadrática sumatoria con valores reales positivos y los cuales se pueden disponer en su orden de decrecimiento:

$$\lambda_1 \geq \lambda_2 \geq \dots, \geq \lambda_i \geq \lambda_{i+1} \geq \dots$$

2. El núcleo se puede representar por una serie monótona convergente, que incluye las funciones propias  $\psi = \{\psi_i\}$  del operador  $\mathcal{A}_\xi$ , de la forma:

$$R_\xi(t, \tau) = \sum_{i=1}^{\infty} \lambda_i \psi_i(t) \psi_i^*(\tau) \quad (3.51)$$

3. Las funciones propias pueden ser ortonormales, de tal forma que se cumpla (1.5), esto es,

$$\langle \mathcal{A}_\xi \{\psi_i\}, \psi_k \rangle = \langle \lambda_i \psi_i, \psi_k \rangle = \lambda_i \delta_{ik} \quad (3.52)$$

En general, en [24] se demuestra que al emplear el operador  $\mathcal{A}_\xi$  para cualquier conjunto base  $\{\phi_i\}$ , es cierto que,

$$\sum_{i=1}^p \langle \mathcal{A}_\xi \{\phi_i\}, \phi_i \rangle \leq \sum_{i=1}^p \langle \mathcal{A}_\xi \{\psi_i\}, \psi_i \rangle = \sum_{i=1}^p \lambda_i$$

La expresión (3.49) para el conjunto base óptimo hallado toma la forma,

$$\bar{\varepsilon}_T^2 = \int_T R_\xi(t, \tau) dt - \sum_{i=1}^p \langle \mathcal{A}_\xi \{\psi_i\}, \psi_i \rangle \quad (3.53)$$

en la cual, al reemplazar (3.51) y (3.52) se obtiene que:

$$\bar{\varepsilon}_T^2 = \sum_{i=1}^{\infty} \lambda_i \langle \psi_i, \psi_i \rangle - \sum_{i=1}^p \lambda_i = \sum_{i=p+1}^{\infty} \lambda_i \quad (3.54)$$

De lo anteriormente expuesto, se puede concluir que el subespacio con dimensión  $p$ , definido en  $L^2(T)$  y que es óptimo para la representación de un proceso aleatorio en  $t \in T$ ,

está dado por las  $p$  funciones propias de la ecuación

$$\int_T R_\xi(t, \tau) \psi_i(\tau) d\tau = \lambda_i \psi_i(t) \quad (3.55)$$

las cuales corresponden a los  $p$  mayores valores de  $\lambda_i$ . Por cierto, debido a que el error cuadrático medio de representación  $\bar{\epsilon}_T^2$  corresponde a la suma residual de los valores propios, entonces el error de aproximación se puede variar aumentando o disminuyendo el índice inferior  $p$  de la respectiva suma en (3.54). Así, en general, aumentando  $T$ , los valores propios también aumentan, luego es necesario una mayor cantidad  $p$  de términos de la descomposición para obtener la precisión deseada.

La descomposición de señales aleatorias que tienen función de correlación continua por la serie (3.16),  $\xi(t) \approx \sum_{i=1}^n \xi_i \psi_i(t)$ ,  $t \in T$ , en la cual el conjunto base óptimo de representación (3.53) son las funciones propias, corresponde a la descomposición *no correlacionada* de Karhunen-Loève (ver ec. (3.16)), en la que los coeficientes (3.17)  $\{\xi_i\}$  son ortogonales:

$$\begin{aligned} E\{\xi_m \xi_l^*\} &= E\{\langle \xi(t), \psi_m(t) \rangle \langle \xi(t), \psi_l(t) \rangle^*\} = \int_T \int_T R_\xi(t, \tau) \psi_m^*(t) \psi_l(t) dt d\tau \\ &= \lambda_l \delta_{ml} \end{aligned}$$

De otra parte, si se tiene un proceso aleatorio con valor medio  $\overline{\xi(t)} = 0$ , entonces los coeficientes, también tendrán valor medio  $\{\overline{\xi_i} = 0, \forall i\}$  y tiene correlación nula (son linealmente independientes).

Aunque la descomposición K-L brinda la menor cantidad de elementos en la representación de procesos aleatorios por medio de la serie (1.4), para un valor dado de error  $\bar{\epsilon}_T^2$ , sin embargo su empleo está limitado principalmente por las siguientes razones: la función de correlación de cualquier proceso aleatorio, en la mayoría de los casos, es desconocida y el procedimiento de solución de (3.55) no es conocido en forma general.

Usualmente, en calidad de funciones base se emplean las funciones ortogonales del tipo Fourier, polinomios de Chebyshev, Lagrange, Funciones de Walsh y Haar, entre otras. No obstante, las funciones base exponenciales de Fourier (1.8),  $\psi_i(t) = e^{j\omega_0 t}$ , son óptimas para la representación de procesos aleatorios estacionarios cíclicos  $\xi(t)$ , para los cuales se cumple,

$$E\{\xi(t + kT)\} = E\{\xi(t)\} = m_{1\xi}(t), \quad R_\xi(t_1 + kT, t_2 + mT) = R_\xi(t_m, t_2)$$

En este caso, los coeficientes de descomposición  $\{\xi_i\}$  determinadas como

$$\xi_k(t) = \frac{1}{T} \int_0^T \xi(t) e^{-jk\omega_0 t} dt \quad (3.56)$$

tienen correlación 0 y su varianza en la expresión (3.20), teniendo en cuenta (3.53) y (3.55),

está dada por el valor:

$$\sigma_{\xi_i}^2 = \int_T R_{\xi}(\tau) \exp(-ji\omega_0\tau) d\tau$$

Las funciones base de Fourier no son óptimas para procesos aleatorios diferentes a los cíclicos estacionarios. Por lo tanto, los coeficientes respectivos de descomposición serán correlacionados. Sin embargo, para procesos ergódicos, cuando  $T \rightarrow \infty$  los coeficientes de Fourier resultan tener correlación nula [15] y la transformada de Fourier se aproxima a la descomposición K-L. Además, se puede demostrar que la relación entre el error de representación y el número mínimo  $p$  de elementos necesarios en la serie (3.16) al emplear la funciones base de Fourier, está dada por [22],

$$\bar{\epsilon}_{T(F)}^2 \approx \frac{2\sigma_{\xi}^2}{\pi^2 p}$$

De otra parte, el sistema de representación de los polinomios de Chebyshev son óptimos en el sentido del criterio de aproximación uniforme de la representación, cuando se cumple que el valor

$$\max \left| \xi^{(k)}(t) - \sum_{n=0}^{N-1} \xi_n T_n(t) \right|$$

es cercano al mínimo posible, que pueda ser obtenido por algún polinomio de aproximación uniforme polinomial [22], donde  $\xi^{(k)}$  es la  $k$  observación del proceso.

El cálculo de los coeficientes  $\xi_k$  para la representación de las señales aleatorias usando los Polinomios de Chebyshev es difícil, debido al valor que toma la función de peso definida en (1.9)  $w(t) = 1/\sqrt{1-t^2}$ . En este sentido es preferible el uso de polinomios con peso constante  $w(t) = 1$ , por ejemplo los polinomios de Legendre, que aunque brindan un valor de aproximación peor que los polinomios de Chebyshev, en cambio ofrecen mayor comodidad en el cálculo de los respectivos coeficientes de representación.

En otro caso, el sistema de funciones de Walsh  $\{\text{wal}_n(t) : n = 0, 1, \dots, 2^m - 1\}$  se considera óptimo en el sentido de representación de señales aleatorias diádico-estacionarias [22], cuya función de correlación cumple la condición  $R(t, \tau) = R(t \oplus \tau)$ . Esta clase de ruido puede ser obtenido si a la entrada del filtro de Walsh se introduce RBG. En otras palabras, la transformada de Walsh de una señal aleatoria diádico-estacionaria se puede identificar con la transformada K-L.

### 3.2.3. Estimación de la densidad espectral de potencia

La DEP  $S_{\xi}(\omega)$ , de una parte, corresponde a la varianza de la descomposición espectral del proceso estacionario, tal y como se muestra en (3.25). Pero de otra parte, la DEP corresponde a la transformada inversa de la función de correlación (3.27). Por lo tanto, la estimación de la DEP se puede realizar mediante ambas formas de representación. Al

analizar la DEP como la densidad de la varianza espectral, la estimación sobre una trayectoria de longitud finita del proceso se hace a través de la descomposición en la serie de Fourier (3.16) de la señal aleatoria para  $\omega = \omega_k$ , esto es,

$$\tilde{S}_\xi(\omega) = \frac{1}{2T} (\alpha_k^2 + \beta_k^2) = \frac{1}{T} \left| \int_0^T \xi(t) e^{-j\omega t} dt \right|^2 \quad (3.57)$$

donde  $\alpha_k = \int_0^T \xi(t) \cos 2\pi kt / T dt$  y  $\beta_k = \int_0^T \xi(t) \sin 2\pi kt / T dt$ .

Debido a la ortogonalidad de la representación y al tomar el valor límite  $T \rightarrow \infty$ , entonces, ambos valores  $\alpha_k$  y  $\beta_k$  no son correlacionados. Además, en los procesos con estructura Gaussiana,  $\alpha_k$  y  $\beta_k$  como se explicó en la sección §3.1.3, también tienen FDP Gaussiana y, por ende, la estimación de  $\tilde{S}_\xi(\omega)$ , haciendo  $T \rightarrow \infty$ , corresponde a la  $\chi^2$ -densidad, (2.31), con dos grados de libertad, con lo cual se demuestra que la media y la varianza de la estimación corresponden a:

$$E \{ S_\xi(\omega_k) \} = S_\xi(\omega_k) \quad (3.58a)$$

$$\sigma_S^2 = S_\xi^2(\omega_k) \quad (3.58b)$$

Mientras, el valor de la media muestra ausencia de sesgo, la varianza no tiende a cero, motivo por la cual, la estimación (3.57) no se considera consistente.

La segunda forma de representación, a partir de (3.27) (empleando la estimación (3.47)), implica la siguiente forma de cálculo para la DEP:

$$\tilde{S}_\xi(\omega) = \int_{-T}^T e^{-j\omega\tau} \tilde{R}_\xi(\tau) d\tau = \int_{-T}^T \left( 1 - \frac{|\tau|}{T} \right) R_\xi e^{-j\omega\tau} d\tau \quad (3.59)$$

Sin embargo, el valor medio de (3.59) muestra un sesgo en la respectiva estimación:

$$E \{ \tilde{S}_\xi(\omega) \} = E \left\{ \int_{-T}^T \left( 1 - \frac{|\tau|}{T} \right) \tilde{R}_\xi e^{-j\omega\tau} d\tau \right\} = \tilde{S}_\xi(\omega) - 2 \int_T^\infty \cos \omega\tau R_\xi(\tau) d\tau$$

Por cuanto [3],

$$\lim_{v \rightarrow \infty} \int_a^b x(\zeta) \cos(v\zeta) d\zeta = \lim_{v \rightarrow \infty} \int_a^b x(\zeta) \sin(v\zeta) d\zeta = 0$$

entonces, para valores pequeño de  $T$ , se tienen distorsiones significativas en el valor de la estimación propuesta.

Sea la observación  $\{ \xi(t), T \in [a, b] \}$  para un proceso estacionario. De ambos métodos de estimación de la DEP, (3.57) y (3.59), se establece la estadística común, denominada

*periodograma* y definida por la relación:

$$s(\lambda; a, b) = \frac{1}{2\pi(b-a)} \left| \int_a^b \xi(t) e^{j t \lambda} dt \right|^2$$

que en el caso discreto, cuando  $\{\xi_k \in T : t_k, k = 1, \dots, n\}$ , toma la forma:

$$s(\lambda; n) = \frac{1}{2\pi n} \left| \sum_{t \in T} \xi(t) e^{j t \lambda} \right|^2 \quad (3.60)$$

Sin embargo, la estadística en (3.60) presenta las mismas limitaciones de inconsistencia dadas en (3.58a) y (3.58b), razón por la cual el periodograma se convierte en otro proceso aleatorio que, para valores grandes de  $n$ , presenta fuertes fluctuaciones de trayectoria (realizaciones con alto grado de disimilaridad). La ausencia de consistencia en las estimaciones propuestas de DEP ha generado la necesidad de empleo de diferentes métodos orientados a aumentar la efectividad de estimación por medio de diversas funciones de peso, que impliquen su alisamiento. En particular, para una observación dada  $\{\xi(t_k) : t_k \in T, k = 1, \dots, n\}$  se emplea la estadística,

$$\tilde{S}_\xi(\lambda) = \int_{-\pi}^{\pi} W_S(\lambda - \nu) s(\nu; n) d\nu \quad (3.61)$$

La función de peso  $W_S(\lambda)$  en frecuencia, (3.61), es similar a la ventana propuesta para la estimación en el tiempo, (3.8), por lo que se denomina *ventana espectral*, que cumple los requerimientos:

- (a). *Localización.*  $W_S(\lambda)$  debe tener un valor máximo en  $\lambda = 0$ ,
- (b). *Ausencia de sesgo.*  $\int_{-\pi}^{\pi} W_S(\lambda) d\lambda = 1$ ,
- (c). *Consistencia.*  $\sigma_S^2 \rightarrow 0$ , cuando  $n \rightarrow \infty$ . En particular, se considera el valor asintótico,

$$\lim_{n \rightarrow \infty} \frac{1}{n} \int_{-\pi}^{\pi} W_S^2(\lambda) d\lambda = 0$$

En la práctica es común el empleo de funciones de peso que se puedan representar en la clase:

$$W_n(\lambda) = 2 \sum_{-n+1}^{n+1} k_n(l) e^{-j \lambda l}$$

donde  $k_n(l) = k(l/m_n)$ , siendo  $\{m_n\}$  una sucesión creciente acotada de valores enteros,

<i>Ventana</i>	<i>Modelo</i>
Transformada finita de Fourier (Estimación de Daniell)	$W_S(\lambda) = \begin{cases} m_n, &  \lambda  \leq \pi/m_n \\ 0, &  \lambda  > \pi/m_n \end{cases}$ $\tilde{S}_\xi(\lambda) = \frac{m_n}{2\pi} \sum_{t=-n+1}^{n-1} \left(1 - \frac{ t }{n}\right) \tilde{R}_\xi(t) \frac{1}{\pi t} \sin \frac{\pi t}{m_n} e^{jt\lambda}$ $k(\nu) = \text{sinc}(\pi\nu)$
Estimación truncada	$W_S(\lambda) = 2 \sin\left(\frac{2m_n+1}{2}\lambda\right) \left(\sin \frac{\lambda}{2}\right)^{-1}$ $\tilde{S}_\xi(\lambda) = \frac{1}{2\pi} \sum_{t=-m_n}^{m_n} \left(1 - \frac{ t }{n}\right) \tilde{R}_\xi(t) e^{jt\lambda}$ $k(\nu) = \begin{cases} 1, &  \nu  \leq 1 \\ 0, &  \nu  > 1 \end{cases}$
Estimación de Bartlett	$W_S(\lambda) = \frac{\sin^2 \frac{m_n \lambda}{2}}{m_n \sin^2 \frac{\lambda}{2}}$ $\tilde{S}_\xi(\lambda) = \frac{1}{2\pi} \sum_{t=-m_n}^{m_n} \left(1 - \frac{ t }{n}\right) \left(1 - \frac{ t }{m_n}\right) \tilde{R}_\xi(t) e^{jt\lambda}$ $k(\nu) = \begin{cases} 1 -  \nu , &  \nu  \leq 1 \\ 0, &  \nu  > 1 \end{cases}$
Estimación de Tukey-Hamming	(ver estimación truncada) $\tilde{S}_\xi(\lambda) = \frac{1}{2} \tilde{S}_{\xi F}(\lambda) + \frac{1}{n} \tilde{S}_{\xi F}\left(\lambda - \frac{\pi}{m_n}\right) + \frac{1}{n} \tilde{S}_{\xi F}\left(\lambda + \frac{\pi}{m_n}\right)$ $k(\nu) = \begin{cases} (1 + \cos \pi\nu)/2, &  \nu  \leq 1 \\ 0, &  \nu  > 1 \end{cases}$

**Tabla 3.1.** Ventanas de estimación de densidad espectral de potencia

tales que  $m_n/n \rightarrow 0$ , cuando  $n \rightarrow \infty$ , donde  $k(x)$  es una función par acotada, que cumple las restricciones:

(a).  $k(0) = 1$ ,  $|k(\nu)| < 1$ ,  $\forall \nu < 1$ ,

(b). Además,  $\int_{-\infty}^{\infty} k^2(\nu) d\nu < \infty$ .

La Tabla 3.1 muestra ejemplos de ventanas espectrales empleadas en la estimación de la DEP. En calidad de estimación de la función de correlación  $\tilde{R}_\xi(t)$  se puede tomar la expresión (3.48).

### 3.2.4. Estimación de parámetros de regresión

Sea una señal aleatoria observada en la forma:

$$\xi(t) = \sum_{k=1}^r \theta_k a_k(t) + \epsilon_0(t) = x(t, \theta_1, \dots, \theta_r) + \epsilon_0(t)$$

donde  $\xi_0(t)$  es un proceso estacionario con valor medio cero,  $\{a_k : k = 1, \dots, r\}$  funciones determinísticas conocidas, asumidas como de carácter lineal,  $\{\theta_k : k = 1, \dots, r\}$  el conjunto de los parámetros desconocidos a estimar, siendo  $r$  el modelo del sistema (valor conocido). El término  $x(\cdot)$  de acuerdo a la clase de tareas de que resuelva, recibe el nombre de *señal útil* (comunicaciones, proceso de señales, etc.) o *tendencia* (en aplicaciones médicas, biológicas, sociológicas, etc.). Mientras, el término  $\epsilon_0(t)$  recibe el nombre de *ruido*.

En general, existen los siguientes métodos de estimación del vector  $\theta = (\theta_0, \theta_1, \dots, \theta_{m-1})$ , que describe los parámetros del modelo de dependencia lineal (2.78).

**Método de mínimos cuadrados.** A partir de la trayectoria  $x(t)$  del proceso estacionario  $\xi(t)$ , la estimación de  $\{\theta_k : k = 1, \dots, r\}$  se realiza minimizando el funcional, derivado del criterio de minimización de la potencia de error (2.63):

$$\int_a^b \left| x(t) - \sum_{k=1}^r \theta_k a_k(t) \right|^2 dt \quad (3.62)$$

Si a su vez, se tiene que  $\{a_k(t) \in L^2(a, b) : k = 1, \dots, r\}$ , entonces, se demuestra que

$$\tilde{\theta}_k = \sum_{l=1}^r c_{kl}^{-1} \int_a^b \overline{a_l(t)} x(t) dt$$

siendo  $c_{kl}^{-1}$  el elemento  $kl$  de la matriz, que corresponde a la inversa de la matriz con elementos:

$$c_{kl} = \int_a^b \overline{a_k(t)} a_l(t) dt$$

Una ventaja del método de los mínimos cuadrados está en que no exige el conocimiento de las propiedades, ni espectrales ni de correlación, del proceso  $\xi(t)$ . Además, no tiene sesgo, y en caso de asumir que el respectivo espectro de potencia de  $S_\epsilon(\lambda)$ , es acotado y positivo definido, se demuestra la consistencia de la estimación de  $\{\theta_k : k = 1, \dots, r\}$ , con la siguiente condición de integrabilidad:

$$\int_{-\infty}^{\infty} |a(t)|^2 dt < \infty$$

**Estimación lineal de mínimo sesgo.** Cuando se conoce la densidad espectral de energía  $F_\epsilon = \mathcal{F}\{\epsilon_0(t)\}$  y, por ende, su función de correlación  $R_\epsilon(\tau)$ , se puede asumir que las funciones  $a_k(t)$ , que forman la regresión base  $x(t)$ , son tales que el proceso  $\xi(t)$  permite su representación espectral en la forma:

$$\xi(t) = \int_{-\infty}^{\infty} e^{jt\lambda} d\zeta(\lambda)$$



para la cual se cumple que

$$d\zeta(\lambda) = \sum_{k=1}^r \theta_k \overline{\alpha_k(\lambda)} dF_\epsilon(\lambda) + d\zeta_0(\lambda)$$

donde  $d\zeta_0(\lambda)$  corresponde al proceso espectral

$$\epsilon_0(t) = \int_{-\infty}^{\infty} e^{jt\lambda} d\zeta_0(\lambda)$$

mientras  $\{\alpha_k(\lambda) : k = 1, \dots, r\}$  son funciones integrables en el sentido cuadrático sobre la medida espectral  $F_\epsilon(\cdot)$ :

$$\int_{-\infty}^{\infty} |\alpha_k(\lambda)|^2 dF_\epsilon(\lambda) < \infty$$

entonces,  $\{\alpha_k \in L^2(F_\epsilon) : k = 1, \dots, r\}$  corresponden a la solución de la ecuación integral de Wiener-Hopf:

$$\int_{-\infty}^{\infty} e^{jt\lambda} \alpha_k(\lambda) dF_\epsilon(\lambda) = a_k(t), \forall k \quad (3.63)$$

Cuando existe la solución de (3.63), la tarea de estimación del conjunto de parámetros de la regresión  $\{\theta_k : k = 1, \dots, r\}$  consiste en la solución del sistema algebraico de ecuaciones lineales.

## Problemas

**Ejercicio en el CP 3.1.** Simular una sucesión aleatoria  $\{\xi[n] : n = 0, \dots, N-1\}$ , para diferentes valores de longitud de trayectoria  $N = 2^m$ ,  $m = 4, \dots, 12$ , con FDP  $\mathcal{N}(0, 1)$ . Verificar de forma cualitativa la normalidad, de acuerdo con la estimación de los valores para los coeficientes de asimetría y exceso.

**Ejercicio en el CP 3.2.** Sea el proceso aleatorio  $\eta(t)$  compuesto por la suma de  $N$  señales estacionarias  $\xi_n(t)$ , estadísticamente independientes, todas con FDP Gaussiana  $\mathcal{N}(0, \sigma_\xi)$ . Hallar la media, varianza y FDP del proceso  $\eta(t)$ . Comprobar experimentalmente, para un valor de significación de  $\alpha = 0.05$  los resultados obtenidos y establecer el comportamiento del error de estimación de los momentos del proceso de acuerdo a la longitud de las trayectorias de las señales  $\xi_n(t)$ .

**Ejercicio en el CP 3.3.** Sea el proceso aleatorio  $\xi(t)$ , que se asume con FDP Gaussiana  $\mathcal{N}(0, \sigma_\xi)$  y función de correlación

$$R_\xi(\tau) = \sigma_\xi^2 \exp(-\alpha |\tau|) \left( \cos \omega \tau + \frac{\alpha}{\omega} \sin \omega |\tau| \right)$$

Hallar la probabilidad  $P\{\dot{\xi} > c\}$  de que el valor de la derivada  $d\xi/dt$  no exceda el valor  $c$ . Simular dependencia entre el valor encontrado de probabilidad encontrado y diferentes valores del coeficiente  $\alpha$ .

**Ejercicio en el CP 3.4.** A la entrada del circuito  $RC$ , mostrado en la Figura 1.3 en régimen de integrador, se tiene RBG  $\xi(t)$  con parámetros  $\mathcal{N}(m_{1\xi}, \sigma_\xi)$ . Hallar el tiempo  $t_c$ , cuando el valor medio del proceso aleatorio  $\eta(t)$  a la salida del integrador alcanza el valor  $c = 0.95m_{1\eta}$ . Hallar la dependencia de  $t_c$  con respecto al valor medio  $m_{1\xi}$  para diferentes valores de la constante de integración  $\tau = RC$ .

**Ejercicio en el CP 3.5.** Realizar el cálculo numérico de la varianza  $\sigma_\eta^2(t)$  de la integral del proceso aleatorio  $\xi(t)$  con función de correlación  $R_\xi(\tau) = \sigma_\xi^2 \exp(-\alpha |\tau|) (1 - \alpha |\tau|)$ , para diferentes valores de la constante de integración  $\alpha$ .

**Ejercicio en el CP 3.6.** A la entrada del circuito  $RC$ , mostrado en el ejemplo (1.3) en régimen de integrador, se tiene la señal aleatoria  $\xi(t)$  con función de correlación y DEP, respectivamente dadas por la expresiones

$$R_\xi(\tau) = \sigma_\xi^2 \exp(-\alpha |\tau|), \quad S_\xi(\omega) = \sigma_\xi^2 \frac{1/\alpha}{1 + (\omega/\alpha)^2}$$

Hallar las correspondientes funciones de correlación  $R_\eta(\tau)$  y DEP  $S_\eta(\omega)$  del proceso  $\eta(t)$  a la salida del integrador, así como su valor de la varianza a la salida del circuito,  $\sigma_\eta^2$ . Realizar la simulación de las expresiones asintóticas de  $R_\eta(\tau)$  para  $\tau_0 \gg \tau$  y  $\tau_0 \ll \tau$ , siendo  $\tau = RC$ . Simular el comportamiento del sistema cuando a la entrada se presenta un tren pulsos cuadrados como el representado en la Figura 3.3.

**Ejercicio en el CP 3.7.** A la entrada de un sistema se tiene una observación de un proceso aleatorio  $\xi(t)$  con FDP diferente a la Gaussiana. El proceso aleatorio de salida es discretizado, de tal forma que puede ser representado en forma aproximada por la suma  $\eta[n] = \sum_k h[k] \xi[n-k]$ . Hallar experimentalmente, para que condiciones la FDP del proceso a la salida puede ser considerada normal.

### 3.3. Estimación espectral no paramétrica

#### 3.3.1. Método de los periodogramas

Dada una sucesión inicial de valores discretos, formada de la trayectoria de señal aleatoria  $\{x[k] : x \in \xi\}$  con longitud  $N$ , a partir de la expresión (3.60), el cálculo del periodograma, mediante técnicas de proceso digital se puede escribir como:

$$s_x[k] = \frac{1}{N} |X[k]|^2 = \frac{1}{N} \left| \sum_{n=0}^{N-1} x[n] e^{-jnk} \right|^2, \quad X[k] = \sum_{n=0}^{N-1} x[n] e^{-jnk} \quad (3.64)$$

Luego, el periodograma se puede calcular empleando la TRF, como se muestra en el procedimiento `Periodograma`. La ec. (3.64) es la forma no modificada de cálculo del periodograma. Sin embargo, como se observa de (3.58b), la operación (3.64) no resulta en una estimación consistente. Con el fin de obtener estimaciones consistentes de la DEP, sobre observaciones discretas de longitud finita, se emplea el suavizado de la estimación (3.64) en frecuencia o tiempo.

Así, la expresión (3.64) se escribe como

$$s_x[k] = \frac{1}{N} \left| \sum_{n=-\infty}^{\infty} w_s[n] x[n] e^{-jnk} \right|^2$$

donde  $w_s[n]$  es la ventana rectangular.

No obstante, como antes se dijo, es posible el empleo de funciones ventana diferentes a la rectangular. El periodograma, calculado mediante una función diferente a la ventana rectangular, se denomina *modificado*, el cual se describe como

$$s_x[w; k] = \frac{1}{NV} \left| \sum_{n=-\infty}^{\infty} w[n] x[n] e^{-jnk} \right|^2$$

donde  $V = \left| \sum_{n=0}^{N-1} w[n] \right|^2 / N$  es un factor de normalización que asegura que la estimación del histograma, asintóticamente, no tenga sesgo. El procedimiento `Periodograma modificado` realiza el periodograma para diferentes ventanas.

```
% Periodograma
function sx = periodogram(x,n1,n2)
    if nargin == 1,
        n1 = 1;
        n2 = length(x);
    end;
    sx = abs(fft(x(n1:n2),1024)).^ 2;
    sx = sx/(n2 - n1 + 1);
    sx(1) = sx(2);
end;
```

```
%Periodograma modificado
function sx = mper(x,win,n1,n2)
    if nargin == 2,
        n1 = 1;
        n2 = length(x);
    end;
    N = n2-n1+1; w = ones(N,1);
    if win==2,
        w = hamming(N);
    elseif win==3, w = hanning(N);
    elseif win==4, w = bartlett(N);
    elseif win==5, w = blackman(N);
    end;
    xw = x(n1:n2).*w/norm(w);
    sx = N*periodogram(xw);
```

Como se explicó para la expresión (3.61), la introducción de una función ventana genera el alisamiento en la frecuencia del histograma, con lo cual, la estimación de la DEP es,

$$E\{s_x[w; k]\} = \frac{1}{2\pi NV} |s[k] * W[k]|^2 \quad (3.65)$$

donde  $W[k]$  es la TF de la función ventana  $w[n]$ . De (3.65), se observa que el nivel de alisamiento en el periodograma depende de la forma de la función  $w[n]$  empleada en el preproceso de la sucesión de entrada. Así, por ejemplo la ventana rectangular tiene el lóbulo principal muy angosto, con relación a la mayoría de ventanas conocidas, lo cual genera menor grado de alisamiento espectral, pero en cambio, esta ventana presenta el mayor nivel de lóbulos laterales, que pueden llevar al enmascaramiento de componentes espectrales de poca amplitud. Cabe anotar, que la estimación mediante el histograma modificado presenta las mismas limitaciones, descritas en (3.58a) y (3.58b), sin embargo, el empleo de las ventanas permite establecer un cierto balance entre la resolución espectral (ancho de banda del lóbulo principal) y la pérdida espectral (área de los lóbulos laterales).

**Promedio de histogramas.** Como consecuencia de la ausencia de sesgo en la operación

$$\lim_{N \rightarrow \infty} E\{s_x[k]\} = S_x[k]$$

se asume que al hallar una estimación consistente para  $E\{s_x[k]\}$ , la misma se mantiene consistente para  $S_x[k]$ . Luego, partiendo del hecho de que la operación de promediado, para un conjunto de observaciones no correladas de la señal aleatoria, conlleva a la estimación consistente del promedio de la variable aleatoria  $E\{\xi\}$ , entonces, se analiza la estimación de la DEP mediante el promedio de los periodogramas o *método de Bartlett*.

Sea  $x_i[n] \in \xi$ ,  $n = 0, \dots, L-1$ ,  $i = 1, 2, \dots, K$ , un conjunto de observaciones no correlacionadas de la señal aleatoria  $\xi(t)$ . Si a la trayectoria  $i$  le corresponde el histograma  $s_x[i, k]$ , entonces el valor medio del conjunto de histogramas será

$$\tilde{S}_x[k] = \frac{1}{K} \sum_{i=1}^K s_x[i, k]$$

cuya esperanza matemática es igual a

$$E\{\tilde{S}_x[k]\} = E\{s_x[i, k]\} = \frac{1}{2\pi} S_x[k] * W_B[k]$$

donde  $W_B(\lambda)$  es la TF de la ventana de Bartlett  $w[n]$ ,  $-L < m < L$ . Por lo tanto, como en el caso del periodograma, el valor medio  $\tilde{S}_x[k]$ , asintóticamente no presenta sesgo. Además, asumiendo la correlación 0 de las trayectorias  $x_i[n]$ , la varianza de la estimación  $\tilde{S}_x[k]$  es:

$$\text{var}\{\tilde{S}_x[k]\} = \frac{1}{K} \sum_{i=1}^K \text{var}\{s_x[i, k]\} \approx \frac{1}{K} S_x^2[k]$$

que tiende a cero, cuando  $K \rightarrow \infty$ . Luego, la estimación  $\tilde{S}_x[k]$  es consistente.

En la práctica, disponer de una cantidad amplia de trayectorias de un mismo proceso es muy difícil, sin embargo, es más frecuente disponer de una trayectoria de longitud suficientemente larga  $N$ . Entonces se puede generar un conjunto de periodogramas obtenidos de la división de la trayectoria inicial en  $K$  segmentos sin traslapo, empleando técnicas de análisis en tiempo corto, cada uno de ellos de longitud  $L$ , de tal manera que  $K < L < N$ .

El procedimiento **Periodograma de Bartlett** ilustra el cálculo de los periodogramas de Bartlett. En general, la estimación promediada de histogramas tiene la forma

$$\tilde{S}_{xB}[k] = \frac{1}{N} \sum_{i=1}^K \left| \sum_{n=0}^{L-1} x[iL + n] e^{-jnk} \right|^2$$

Se puede considerar que la estabilidad de la estimación  $\tilde{S}_{xB}[k]$  mejora, al disminuir la cantidad de segmentos de análisis de tiempo corto. Dado el valor  $N = KL$  se observa la relación de compromiso entre una alta resolución (cuando se tiene el máximo valor de  $L$ ) y de mínima varianza de la estimación (para el máximo valor posible de  $K$ ).

#### Promedio de histograma modificado.

El análisis de tiempo corto mediante segmentos sin traslapo puede generar discontinuidades en los valores contiguos de estimación. Por esta razón, es preferible el empleo de segmentos de traslapo que aseguren la suavidad en la estimación. De otra parte, el empleo de funciones ventana puede atenuar el efecto de la pérdida espectral, así como disminuir el sesgo de la estimación, sin embargo, a costa de una pérdida en la resolución espectral. Así, sea el valor de traslapo  $D$ , para segmentos de análisis de tiempo corto con longitud  $L$ , luego la trayectoria  $i$  tiene la forma

$$x[i, n] = x[n + iD], \quad n = 0, 1, \dots, L - 1$$

Al tomar  $K$  segmentos que cubran la totalidad del registro de longitud  $N$ , entonces, se tiene que  $N = L + D(K - 1)$ , por lo que la estimación del periodograma de Welch se escribe

```
%Periodograma de Bartlett
function sx = bart(x,nsect)
L = floor(length(x)/nsect);
sx = 0; n1 = 1;
for i=1:nsect,
    pe = periodogram(x(n1:n1+L-1));
    sx = sx + pe/nsect;
    n1 = n1 + L;
end;
```

```
%Periodograma de Welch
function sx = welch(x,L,over,win)
if (over>=1) | (over<0),
    error('Valor traslapo inválido');
end;
n1 = 1; n0 = (1-over)*L;
nsect = 1 + floor((length(x)-L)/n0);
sx = 0;
for i=1:nsect,
    mp = mper(x,win,n1,n1+L-1)/nsect;
    sx = sx + mp;
    n1 = n1 + n0;
end;
```

en la forma:

$$\tilde{S}_{xW}[w; k] = \frac{1}{KLV} \sum_{i=0}^{K-1} \left| \sum_{n=0}^{L-1} w[n] x[n + Di] e^{-jnk} \right|^2$$

que en términos del histograma modificado es igual a:

$$\tilde{S}_{xW}[w; k] = \frac{1}{K} \sum_{i=0}^{K-1} s[w; i, k]$$

con lo cual, el valor esperado del periodograma de Welch resulta en:

$$E \left\{ \tilde{S}_{xW}[k] \right\} = E \left\{ s_x[w; k] \right\} = \frac{1}{2\pi LV} S_x[k] |W[k, L]|^2$$

donde  $W[k, L]$  es la TDF de la función ventana con longitud  $L$ , usada en la estimación del periodograma de Welch. Por cuanto, el factor de normalización es

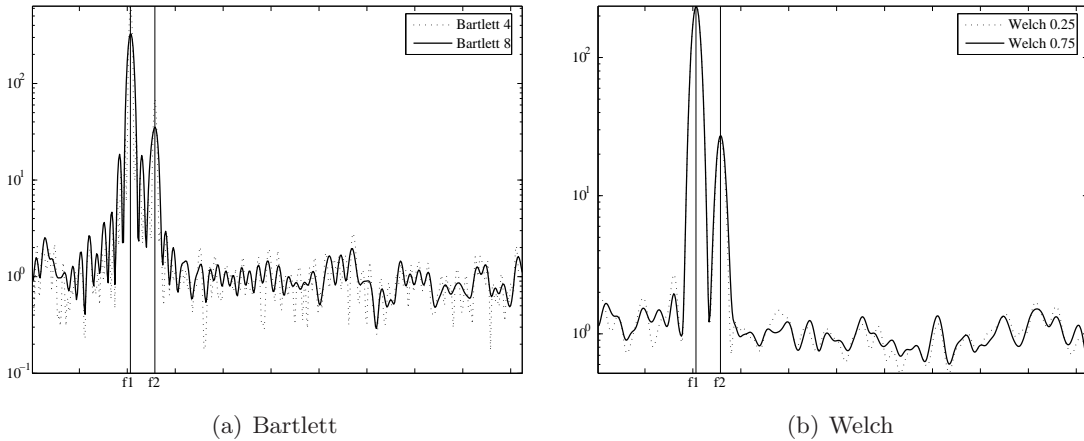
$$V = \left| \sum_{n=0}^{L-1} w[n] \right|^2 / L$$

entonces, el periodograma de Welch también conlleva a una estimación sin sesgo.

**Ejemplo 3.10.** Sea la señal compuesta de dos armónicos y ruido blanco Gaussiano

$$x[n] = a_1 \sin(nk_1) + a_2 \sin(nk_2) + w[n]$$

dados los valores,  $k_1 = 0.2\pi$ ,  $k_2 = 0.25\pi$ ,  $a_1 = \sqrt{10}$ ,  $a_2 = 1$ . Hallar la estimación por Bartlett de la DEP.



**Figura 3.7.** Ejemplo de los periodogramas

En la Figura 3.7(a) se muestran los resultados obtenidos asumiendo los valores  $N = 512$  y

$L = 64$  para dos casos de análisis  $K = 4$  y  $K = 8$ . Como se observa, aunque la varianza de la estimación disminuye, al aumentar la longitud del segmento de tiempo corto  $K$ , empeora la resolución, condicionada por el aumento del lóbulo principal.

**Ejemplo 3.11.** En la Figura 3.7(b) se analiza el efecto del traslapo del ejemplo 3.10, dados los valores 50 y 75 %, en la varianza y resolución espectral del periodograma de Welch.

El cálculo de la varianza en la estimación de Welch es más complejo, en la medida, en que no se puede asegurar la ausencia de correlación entre las trayectorias con traslapo, caso en el cual debe aumentar la dispersión de los valores de la estimación. Por lo tanto, el empleo del traslapo aumenta la cantidad de segmentos de análisis corto, que en un principio debe disminuir la varianza del histograma. No obstante, al aumentar el valor del traslapo, aumenta el costo computacional en un valor proporcional a  $K$ . Además, el consiguiente aumento del traslapo aumenta el valor de correlación entre los segmentos que puede neutralizar el efecto de aumento en la cantidad  $K$ . En la práctica, el valor compromiso de traslapo se escoge entre un 50 y 75 % [25].

**Suavizado de periodogramas en frecuencia.** A partir de la expresión (3.59), se tiene que el periodograma se calcula empleando la TF de la función de correlación propia, en particular:

$$\tilde{S}_x[k] = \sum_{m=-(N-1)}^{N-1} r_x[m] \exp[-jmk] \quad (3.66)$$

donde

$$r_x[m] = \begin{cases} \frac{1}{N} \sum_{n=0}^{N-m-1} x[n+m]x^*[n], & 0 \leq m \leq N-1 \\ \frac{1}{N} \sum_{n=-1}^{N-|m|-1} x^*[n+|m|]x[n], & -(N-1) \leq m \leq -1 \end{cases} \quad (3.67)$$

se define como el *correlograma* de la sucesión aleatoria  $x[n]$  con longitud  $N$ .

De la expresión (3.66), se observa que se emplea un número diferente de operaciones dependiendo del intervalo de correlación. Así, en los extremos se tiene una sola operación. Cuando el intervalo de análisis es  $N-1$ , entonces  $r_x[N-1] = x[N-1]x^*[0]/N$ . Luego, la varianza de la estimación de la función de correlación tendrá valores muy altos en la medida en que crece hacia el valor extremo  $N$  del intervalo de correlación.

La estimación del correlograma, (3.67), implica el empleo de una ventana rectangular, no obstante, se puede variar el tipo de función ventana,  $w_r$ , tal que compense el efecto de las estimaciones no confiables en los extremos del intervalo de correlación, aunque esto implique la disminución del intervalo de proceso útil, y de esta forma, la disminución de la resolución. Además, se aumenta el fenómeno de pérdida espectral, debido a la aparición de lóbulos laterales, con lo cual la estimación tiende a presentar sesgo. Con el objeto de disminuir la pérdida espectral, se puede emplear el suavizado del periodograma, mediante

su convolución con una ventana espectral adecuada o *método de Blackman-Tukey*, que corresponde a la generalización del método de estimación por correlograma para la función de correlación propia, sopesado por una ventana. En este caso, la estimación de la DEP se puede describir como:

$$\tilde{S}_{xBT}[k] = \sum_{m=-M}^M w_r[m] r_x[m] e^{-jmk}$$

que en la frecuencia tiene la forma

$$\tilde{S}_{xBT}[k] = \frac{1}{2\pi} s_x[k] * W[k] \quad (3.68)$$

De (3.68), se ve que la estimación de Blackman-Tukey realiza el suavizado del periodograma mediante su convolución con la TDF de la ventana de correlación  $w_r[m]$ , la cual en la estimación de los correlogramas, debe cumplir las condiciones:

1.  $0 \leq w_r[m] \leq w_r[0] = 1$
2.  $w_r[-m] = w_r[m]$
3.  $w_r[m] = 0, \forall |m| > M, M \leq N - 1$

De la expresión (3.68), se observa que es suficiente, mas no necesaria, la condición de que el periodograma tenga un valor semidefinido positivo, que corresponde a la restricción  $W(\lambda) \geq 0$ , con  $-\pi \leq \lambda \leq \pi$ .

La ventana triangular de Bartlett cumple esta condición, mas no ocurre con las ventanas de Hamming y rectangulares, comunes en el proceso de señales. En la Figura 3.8 se muestra un ejemplo del cálculo del periodograma mediante el método de Blackman-Tukey, que presenta una varianza mucho menor, aunque empeore su resolución de frecuencia.

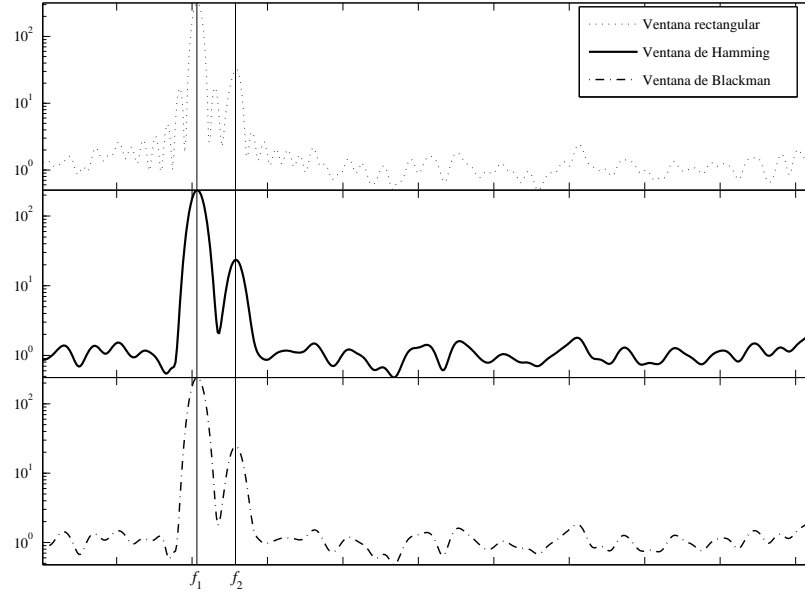
En general, se considera que la estimación de la DEP mediante cualquiera de los métodos de histograma, antes analizados, debe cumplir un compromiso entre la resolución espectral y la varianza de la estimación, por esto, la comparación de los mismos se realiza empleando las siguientes características estadísticas [25]:

$$(a). \nu = \frac{\text{var}\{\tilde{S}_x(\lambda)\}}{E^2\{\tilde{S}_x(\lambda)\}}$$

el cual corresponde al nivel de dispersión de la estimación y caracteriza la estabilidad estadística de la estimación.

```
% Periodograma de Blackman-Tukey
function sx = per_smooth(x,win,M,n1,n2)
if nargin==3, n1=1;
    n2 = length(x);
end;
R = cov(x(n1:n2),M);
r = fliplr(R(1,2:M),
R(1,1),R(1,2:M));
M = 2*M-1;
if win==2 w = hamming(N);
    elseif win==3
        w = hanning(N);
    elseif win==4 w = bartlett(N);
    elseif win==5 w = blackman(N);
end;
r = r'.*w;
sx = abs(fft(r,1024));
sx(1) = sx(2);
```





**Figura 3.8.** Ejemplo de periodograma de Blackman–Tukey

(b).  $\mu = \nu \Delta f$

donde  $\Delta f$  es la capacidad de resolución del método. El parámetro  $\mu$  permite determinar la resolución del DEP obtenida a partir de una trayectoria de longitud finita de la señal aleatoria, por cierto, a menor  $\mu$ , mayor calidad de la estimación.

Método	Estabilidad $\nu$	Resolución $\Delta f$	Índice $\mu = \nu \Delta f$
Periodograma	1	$0.89(2\pi/N)$	$0.89(2\pi/N)$
Bartlett	$1/K$	$0.89(2\pi/N)$	$0.89(2\pi/N)$
Welch	$(9/8)(1/K)$	$1.282(\pi/L)$	$0.72(2\pi/N)$
Blackman-Tukey	$(2/3)(M/N)$	$0.64(2\pi/M)$	$0.43(2\pi/N)$

**Tabla 3.2.** Comparación de los métodos de periodograma de estimación de DEP

En la Tabla 3.2 [25], se muestran las características estadísticas para los métodos del periodograma, descritos anteriormente. Se observa que cada uno de los métodos tiene más o menos el mismo valor compromiso de estabilidad y resolución, el cual es inversamente proporcional a la longitud  $N$  de la sucesión de valores discretos de la señal aleatoria. En general, sin importar que métodos se tengan, que ofrezcan mejor resolución o menor dispersión en la estimación, se cumple que el compromiso entre la resolución espectral y la estabilidad básicamente depende de la longitud,  $N$ , y de los valores discretos disponibles para el análisis de la sucesión aleatoria  $x[n]$ .

### 3.3.2. Algoritmo de cálculo del método de periodograma

Sean conocidos, para el proceso aleatorio en análisis, los valores del intervalo de discretización  $\Delta t$  y la resolución espectral necesaria  $\Delta f$  de análisis. Entonces, el intervalo de observación  $T$  y la respectiva cantidad  $N$  de valores discretizados de la señal se relacionan por la expresión [26]:

$$T = \frac{k_w}{\Delta f}, \quad N = \lfloor \frac{T}{\Delta t} \rfloor \quad (3.69)$$

donde  $k_w$  es un coeficiente que se determina de acuerdo al tipo de ventana empleada en la estimación.

El algoritmo del método de periodogramas se divide en dos etapas:

- el preproceso de la señal en el intervalo de observación,
- el promedio de los resultados durante varios intervalos de aproximación, con el fin de disminuir la dispersión de la estimación.

La primera etapa comprende los siguientes procedimientos:

1. Cálculo del valor  $N$ , a partir de (3.69). Por cuanto la TF se calcula mediante el algoritmo TRF, entonces, cuando  $N \neq 2^m$ ,  $m \in \mathbb{Z}$ , en cada una de las trayectorias, se realiza el relleno de ceros hasta el primer entero  $m$ , para el cual se cumpla la igualdad.
2. Selección de la función ventana  $w_s[n]$  y cálculo de la TRF, para cada una de las  $l$  trayectorias,  $x \in \xi$  disponibles  $\{x[n, k] : n = 0, \dots, N-1; k = l, \dots, L\}$ , que son alisadas por la respectiva ventana:

$$X[k, l] = \sum_{n=0}^{N-1} x[n, l] w_s[n] \exp(-j2\pi kn/N) \quad (3.70)$$

3. Cálculo del periodograma  $s[l, \lambda, N]$ , definido en (3.60), para cada una de las trayectorias  $l$ :

$$s[l, k, N] = \frac{|X[k, l]|^2}{\sum_{n=0}^{N-1} w_s[n]} \quad (3.71)$$

4. Dado un criterio de convergencia, si el valor  $s[l, k, N]$  no lo cumple, entonces, se deben repetir los pasos 1 y 2, sobremuestreando la sucesión de cada trayectoria en 2, 4, 8, ... veces, mientras no se disminuya adecuadamente la incertidumbre en la estimación del periodograma.

Una forma alterna de disminuir la incertidumbre está en realizar el procedimiento de rellenos de ceros aumentando también la longitud  $N$  de cada trayectoria en

$2, 4, 8, \dots, 2^m$  veces. En este caso, el arreglo de entrada de la TRF se modifica en la forma:

$$X[m; k, l] = \sum_{n=0}^{2^m N - 1} x[m; n, l] w_s[n] \exp\left(-j \frac{2\pi k n}{N 2^m}\right) \quad (3.72)$$

donde

$$x[m; n, l] = \begin{cases} x[n, l], & 0 \leq n \leq N - 1 \\ 0, & n > N - 1 \end{cases}$$

Al comparar las expresiones (3.70) y (3.72), se observa que  $X[m; k, l] = X[2^m k, l]$ , luego, la operación de compresión de escala de tiempo expande la envolvente espectral, pero no le cambia de forma. En la práctica, el relleno de ceros conlleva a la aparición de componentes adicionales entre los armónicos originales.

La segunda etapa comprende los siguientes procedimientos:

1. Selección del factor de traslapo  $\Delta$  entre los intervalos contiguos de observación. Usualmente, se toma un valor entre  $\Delta \in [0.5, 0.75]$ .
2. Cálculo del número total de intervalos de observación,  $N_i$ :

$$N_i = \lfloor ((N_T - \Delta N)) / (N - \Delta N) \rfloor$$

donde  $N_T$  es la longitud total del registro de la señal aleatoria en análisis.

3. Estimación de la DEP promediada:

$$\tilde{S}[k] = \frac{1}{N_i} \sum_{l=1}^{N_T} s[l, k, N] \quad (3.73)$$

4. Cálculo del coeficiente  $k_\sigma$ , que muestra la disminución en la varianza de la estimación de la DEP, debido a la operación de promediado, en cada uno de los intervalos de observación.

$$k_\sigma = \begin{cases} \left( \frac{1}{N_i} (1 + 2c^2(0.5)) - \frac{2}{N_i^2} c^2(0.5) \right)^{-1}, & \Delta = 0.5 \\ \left( \frac{1}{N_i} (1 + 2c^2(.75) + 2c^2(0.5)) - \frac{2}{N_i^2} (c^2(0.75) + 2c^2(0.5)) \right)^{-1}, & \Delta = 0.75 \end{cases}$$

donde los valores de  $c(\Delta)$  se determinan de acuerdo al tipo de ventana  $w_s[n]$  empleada en la estimación espectral [27].

### 3.3.3. Ventanas de estimación espectral

La multiplicación de los valores discretizados de la señal aleatoria  $x[n]$  por la función ventana,  $w_s[n]$  corresponde a la convolución en frecuencia de los respectivos espectros.

Entre las principales propiedades y características de las funciones ventana, empleadas en la estimación espectral, están las siguientes:

- Simetría,  $w[n] = w[N - n]$ ,  $n = 1, \dots, N - 1$ .
- Ancho de banda equivalente de ruido, que se determina como

$$\Delta f_R(N) = \frac{N \sum_{n=0}^{N-1} w^2[n]}{\left( \sum_{n=0}^{N-1} w[n] \right)^2}$$

a menor valor de  $\Delta f_R$ , menor es la potencia del ruido y, por lo tanto, menor distorsión que se puede causar sobre el proceso en análisis.

**Ejemplo 3.12.** Hallar el ancho de banda equivalente de ruido por la ventana triangular, cuando  $N = 2K$ , asumiendo que  $w[0] = 0$ .

La ventana triangular se determina como

$$w[n] = \begin{cases} n/K, & n = 1, \dots, K \\ (2K - n)/K, & n = K + 1, \dots, 2K - 1 \end{cases}$$

Entonces,  $\Delta f_R(N) = 2(2K^2 + 1)/(3K^2)$ , por cuanto en la práctica,  $N \gg 1$ , en lugar de  $\Delta f_R(N)$  se emplea su valor asintótico, cuando  $N \rightarrow \infty$ , esto es,  $\Delta f_R = \lim_{N \rightarrow \infty} \Delta f_R(N)$ . Particularmente, en el caso de la ventana triangular  $\Delta f_R = 1.33$ .

- Ancho de banda del lóbulo principal  $\Delta f_{LP}(N, g)$  de la magnitud del espectro, que típicamente se determina para un nivel fijo de atenuación  $g$ , dado en  $dB$ , con relación al valor máximo del módulo espectral,  $\Delta f_{LP}(N, g) = \gamma(N)N/\pi$ , donde  $\gamma(N)$  es la menor de todas las posibles raíces por módulo absoluto, obtenidas de la ecuación,  $|W(N)| = \max |W(N)|/10^{g/20}$ .

En la práctica, es usual el empleo del valor asintótico, dado el nivel  $g$ , para el ancho de banda del lóbulo principal

$$\Delta f_{LP}(g) = \lim_{N \rightarrow \infty} (\gamma(N)N/\pi)$$

Se considera que el valor  $\Delta f_{LP}(g)$  caracteriza la resolución del algoritmo de la TRF, para una función ventana dada, en particular, se asume que [28]

$$\kappa_o = \Delta f_{LP}(g)$$

- Ganancia coherente

$$G_c(N) = \frac{\sum_{n=0}^{N-1} w[n]}{\max |W(N)|}$$

con valor asintótico  $G_c = \lim_{N \rightarrow \infty} G_c(N)$ , que corresponde a la amplificación relativa del armónico de una señal, cuya frecuencia coincide con una de las frecuencias del conjunto base de la TRF.

- Máximo nivel de cualquiera de los lóbulos laterales,  $m_{LL}$ , obtenido para el método de TRF de la función ventana, medido en  $dB$  con relación al valor máximo del lóbulo principal. En este sentido, también se considera la velocidad de caída  $v_{LL}$  de los lóbulos laterales, medida en  $dB$  por octava (o década), que muestra qué tan rápido decrece la energía contenida en los lóbulos laterales. A mayor velocidad de caída, menor pérdida espectral presenta la ventana.
- Modulación de amplitud parásita,  $a_P$ , que caracteriza la amplitud relativa del armónico de la señal, después de su proceso mediante la función ventana y cálculo de la TRF, en el peor de los casos, cuando la frecuencia de la señal se encuentra exactamente en la mitad de un par de frecuencias base de la TRF.

El valor  $a_P$  se mide en  $dB$  y se define como

$$a_P = 20 \log |W(\omega)| / \max(W(\omega))$$

Los siguientes factores confluyen en la elección del tipo de función ventana:

- aplicación concreta del proceso de señales
- exigencias de costo computacional (recursos de proceso y tiempo de cómputo) en la solución de la aplicación

**Ejemplo 3.13.** Sea la sucesión

$$x[n] = \sum_k a_k \sin(2\pi nk/N + \phi_k), \quad n = 0, \dots, N-1 \quad (3.74)$$

donde  $a_k$  y  $\phi_k$  son los valores desconocidos para la amplitud y fase de los armónicos, respectivamente, los cuales coinciden en la frecuencia base de la TRF.

En este caso, el cálculo de  $a_k$  y  $\phi_k$  se puede realizar, simplemente empleando la función ventana rectangular. En particular, de (3.70) se tiene que

$$a_k = |X[k]|, \quad \phi_k = \arg(X[k])$$

Sin necesidad de recurrir al procedimiento de promedio, (3.73), en la medida en que no hay incertidumbre de medida.

Si en la expresión (3.74), asumiendo que se tiene un solo armónico, se agrega la perturbación  $\{r[n]\}$ , en forma de RBG, entonces para la estimación de los valores de  $a_k$  y  $\phi_k$  es necesario procesar la sucesión inicial mediante una función ventana, que provea el menor valor de  $\Delta f_R$ , por ejemplo, la ventana rectangular, y luego, se realiza el procedimiento de promediado (3.73).

**Ejemplo 3.14.** Sea la sucesión

$$x[n] = a_1 \sin(n\omega_1 + \phi_1) + a_2 \sin(n\omega_2 + \phi_2) \quad (3.75)$$

donde  $\omega_{11} \leq \omega_1 \leq \omega_{12}$ ,  $\omega_{21} \leq \omega_2 \leq \omega_{22}$ , siendo  $\omega_{ij}$ , las frecuencias base dadas inicialmente de la TRF. Hallar los valores de  $\omega_1$ ,  $\omega_2$ ,  $a_1$  y  $a_2$ , conocido que  $a_1 \approx a_2$ .

En este caso, el preproceso de la sucesión inicial se puede realizar mediante la función ventana rectangular. Asumiendo en (3.69)  $\Delta f = (\omega_{21} - \omega_{12})/2\pi$  y  $\kappa_o = \Delta_{LP}(g)$ , entonces se puede determinar el intervalo de observación  $T$ , con lo que conocido el intervalo de discretización  $\Delta t$  se puede hallar la longitud  $N_i$  de observación.

Sin embargo, el cálculo de la DEP, en correspondencia con (3.70) y (3.71) para  $k = 0, \dots, N-1$ , puede tener lugar una indeterminación que haga imposible estimar cada armónico debido a la presencia de varios valores iguales de  $\hat{S}(k)$ .

La forma directa para disminuir esta incertidumbre consisten en aumentar la longitud de la sucesión inicial agregando  $N(2^m - 1)$  ceros, con lo cual el espectro  $X[k]$  se calcula por la expresión (3.72). Inicialmente se prueba con  $m = 1$ . En caso de que la incertidumbre no permita aún el cálculo confiable de los armónicos, se aumenta sucesivamente  $m = 2, 3, \dots$  hasta que se obtenga un valor adecuado del armónico.

Debido a que la sucesión inicial se considera que no está perturbada, entonces, no hay necesidad de la operación de promedio (3.73).

## Problemas

**Problema 3.9.** Demostrar que la estimación de Bartlett para la DEP no presenta sesgo.

**Problema 3.10.** Demostrar que la función  $k(\nu)$  de la estimación de Tukey-Hamming para la DEP es cuadrado integrable,  $L^2(\mathbb{R})$ .

**Problema 3.11.** Demostrar que  $E\{\tilde{S}_x[k]\} = \frac{1}{2\pi} S_x[k] * W_B[k]$ .

**Problema 3.12.** Repetir el ejemplo 3.10 para la señal  $x[n] = 2 \sin(0.4\pi n) + 2 \cos(4.5\pi n + 0.5) + w[n]$ , con RBG con media 0 y varianza 0.2.

**Problema 3.13.** Demostrar que  $\tilde{S}_{xBT}[k] = \frac{1}{2\pi} s_x[k] * W[k]$ .

**Problema 3.14.** Considérese una sucesión aleatoria de longitud  $N$ , la cual ha de ser dividida en  $K$  segmentos, cada uno con  $M = N/K$  puntos. Si se conoce que la DEP tiene dos picos separados a una distancia de  $2[\text{rad/s}]$ , cual debe ser el mínimo valor de  $M$  para poder detectar correctamente la presencia de ambos picos?

**Problema 3.15.** Sea la sucesión aleatoria, cuya función de correlación propia se da por la expresión  $R[k] = 0.8^{|k|}$ ,  $k = 0, \pm 1, \pm 2, \dots$ . Si se emplean 10 valores discretos para estimar la función de correlación, mediante el correlogramas (3.67). Hallar el sesgo de la estimación para todo los  $k$ .

### 3.4. Estimación espectral paramétrica

En los métodos paramétricos, la señal aleatoria medida se analiza como la salida de un sistema lineal e invariante en el tiempo, al cual se aplica una entrada con características de aleatoriedad definidas a priori. En la práctica, el modelo más usado corresponde al excitado por RBG y con función de transferencia racional del sistema. Los procesos de salida de los modelos de esta clase tienen DEP que se describe completamente, mediante los coeficientes de la respectiva función de transferencia y la varianza del RBG de entrada.

En esencia, los métodos paramétricos incluyen alguna información sobre los procesos analizados, que tiene en cuenta sus propiedades y su naturaleza interna; información que se desprecia en los métodos no paramétricos. Como resultado, los métodos paramétricos permiten obtener estimaciones con mayor precisión y mayor resolución.

La estimación espectral paramétrica consta de tres etapas:

1. Selección del modelo paramétrico de la señal aleatoria analizada
2. Estimación de los parámetros del modelo seleccionado de la señal aleatoria de acuerdo a las trayectorias conocidas
3. Cálculo de la DEP mediante la sustitución en el modelo teórico de los parámetros estimados

#### 3.4.1. Reacción de un sistema lineal a una señal estacionaria

Sea la entrada  $x[n]$  una sucesión estacionaria con valor medio 0, entonces  $y[n]$  representa la salida del sistema lineal e invariante en el tiempo, la cual también es estacionaria y se describe como

$$y[n] = \mathcal{H}\{x[n]\}$$

La función de correlación propia de la respuesta se determina por la convolución [25]:

$$r_y[m] = r_x[m] * \left( \sum_{k=-\infty}^{\infty} h[k+m]h^*[k] \right)$$

que empleando la transformada  $\mathcal{Z}$  se obtiene

$$S_y(z) = S_x H(z) H^*(1/z^*) \quad (3.76)$$

donde  $S_x(z) = \mathcal{Z}\{r_x[m]\}$ ,  $S_y(z) = \mathcal{Z}\{r_y[m]\}$ ,  $H(z) = \mathcal{Z}\{h[m]\}$ .

Sea la entrada RBG, con valor medio igual a 0 y potencia  $\sigma_\eta^2$ , tal que

$$S_x(z) = \mathcal{Z}\{\sigma_\eta^2 \delta[m]\} = \sigma_\eta^2$$

entonces, la salida y la entrada del respectivo sistema lineal e invariante en el tiempo tiene

la forma:

$$S_y(z) = \sigma_\eta^2 H(z) H^*(1/z^*)$$

La última ecuación muestra la relación de proporcionalidad que hay entre la función de transferencia del filtro del modelo y la DEP de señal aleatoria analizada.

### 3.4.2. Modelos paramétricos de señales aleatorias

La descripción paramétrica de las estadísticas de segundo orden también aplica a las sucesiones de discretización de las señales aleatorias. En este sentido, el modelo es adecuado para la aproximación de señales aleatorias discretas, se describe mediante la señal de salida de un sistema lineal e invariante en el tiempo, expresada por la ecuación de diferencias en coeficientes complejos

$$x[n] = - \sum_{k=1}^p a_k x[n-k] + \sum_{k=0}^q b_k u[n-k] \quad (3.77a)$$

$$= \sum_{k=0}^{\infty} h[k] x[n-k] \quad (3.77b)$$

donde  $u[n]$  es la sucesión de excitación o la entrada del sistema,  $x[n]$  es la sucesión a la salida del filtro causal ( $h[k] = 0, \forall k < 0$ ), el cual conforma los datos observados. Se asume que el sistema lineal tiene función de transferencia racional,  $H(z) = B(z)/A(z)$ , para la cual los polinomios se determinan como

$$\begin{aligned} A(z) &= 1 + \sum_{k=1}^p a_k z^{-k} \\ B(z) &= 1 + \sum_{k=0}^q b_k z^{-k}, \quad b_0 = 1 \\ H(z) &= 1 + \sum_{k=1}^{\infty} h_k z^{-k} \end{aligned}$$

Se asume además que los ceros de los polinomios  $A(z)$  y  $B(z)$  se encuentran ubicados dentro del círculo unitario de convergencia del plano  $\mathcal{Z}$ , con el fin de garantizar la estabilidad del filtro.

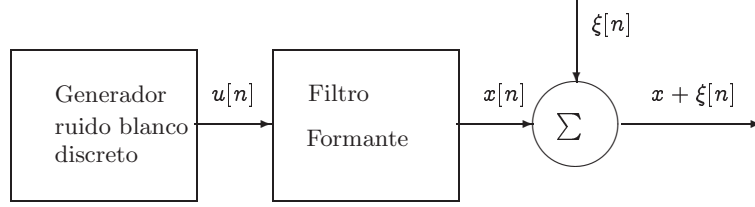
En concordancia con (3.76) los espectros de entrada y salida en un sistema lineal e invariante en el tiempo se relacionan como:

$$S_x(z) = S_u(z) H(z) H^*(1/z^*) = S_u(z) \frac{B(z) B^*(1/z^*)}{A(z) A^*(1/z^*)} \quad (3.78)$$

La sucesión de entrada  $u[n]$ , usualmente, no es observable y, por lo tanto, no se puede emplear en el análisis espectral. Los modelos paramétricos, como se muestra en la Figura



3.9 emplean en calidad de señal de excitación una sucesión de RBG con valor medio cero y varianza  $\sigma^2$ , entonces,  $S_u(z) = \sigma^2$ .



**Figura 3.9.** Diagrama de formación de un modelo paramétrico de un proceso aleatorio

En general, las propiedades de la señal aleatoria modelada, dependen de la estructura y valores de los parámetros del filtro formante, además de las propiedades de la señal de entrada.

En caso de ser necesario el modelado de datos de medición en forma de una mezcla aditiva del modelo paramétrico más alguna perturbación  $\eta[n]$ , entonces, la perturbación se agrega a la salida del filtro formante, como se muestra en la Figura 3.9.

En la estimación de la DEP, se emplean modelos excitados por RBG los cuales se clasifican en tres clases

1. modelos de procesos autorregresivos (AR)
2. modelos de procesos de media deslizante (MA)
3. modelos de procesos autorregresivos y media deslizante (ARMA)

Las diferentes clases, que se muestran en la Tabla 3.3, se diferencian por el tipo de función de transferencia discreta del filtro formante y, en consecuencia, por el tipo de ecuación lineal de diferencia que describe la sucesión de salida aleatoria.

<i>Modelo paramétrico</i>	<i>Ecuación iterativa</i>	<i>Función de transferencia</i>
AR	$x[n] = - \sum_{k=1}^p x[n-k] + u[n]$	$\frac{1}{1 + \sum_{k=1}^p a_k z^{-k}}$
MA	$x[n] = u[n] + \sum_{k=1}^q b_k u[n-k]$	$1 + \sum_{k=1}^q b_k z^{-k}$
ARMA	$x[n] = \sum_{k=1}^p a_k x[n-k] + u[n] + \sum_{k=1}^q b_k u[n-k]$	$\frac{1 + \sum_{k=1}^q b_k z^{-k}}{1 + \sum_{k=1}^p a_k z^{-k}}$

**Tabla 3.3.** Características de los modelos paramétricos

**Modelos ARMA.** Se describen por la ecuación de diferencias generalizada (3.77a), en la cual como se observa de la Tabla 3.3, se tienen polos y ceros. La DEP para un modelo ARMA se obtiene al sustituir en (3.78),  $z = e^{j\Omega}$ :

$$S(\Omega) = \sigma^2 \left| \frac{B(\Omega)}{A(\Omega)} \right|^2 = \frac{\sigma^2 \mathbf{e}_q^H(\Omega) \mathbf{b} \mathbf{b}^H \mathbf{e}_q^H(\Omega)}{\mathbf{e}_p^H(\Omega) \mathbf{a} \mathbf{a}^H \mathbf{e}_p^H(\Omega)} \quad (3.79)$$

donde los vectores columna de los exponentes complejos  $\mathbf{e}_p$  y  $\mathbf{e}_q$ , mas los coeficientes  $\mathbf{a}$  y  $\mathbf{b}$  son de la forma

$$\begin{aligned} \mathbf{e}_p^\top &= (1, e^{j\Omega}, \dots, e^{jp\Omega}), & \mathbf{e}_q^\top &= (1, e^{j\Omega}, \dots, e^{jq\Omega}) \\ \mathbf{a}^\top &= (1, a, \dots, a_p), & \mathbf{b}^\top &= (1, b, \dots, b_q) \end{aligned}$$

Es anotar que la DEP se calcula en el rango de frecuencias normalizadas  $\Omega \in [-\pi, \pi]$ .

De la expresión (3.79), es claro que el modelo ARMA se caracteriza por los parámetros  $a$ ,  $b$  y la varianza del ruido blanco  $\sigma^2$ .

**Modelo MA.** De acuerdo con la función de transferencia mostrada en la Tabla 3.3, la respuesta a impulso en este caso es

$$x[k] = \begin{cases} b_k, & k = 0, 1, \dots, q; \\ 0, & \text{en otro caso.} \end{cases}$$

que tiene longitud finita y, por lo tanto, en concordancia con la expresión (1.57), los filtros formantes de los modelos MA pertenecen a la clase de los no recursivos. La DEP del proceso se calcula haciendo en (3.79)  $p = 0$ , de acuerdo a la expresión:

$$S(\Omega) = \sigma^2 |B(\Omega)|^2 = \sigma^2 \mathbf{e}_q^H(\Omega) \mathbf{b} \mathbf{b}^H \mathbf{e}_q^H(\Omega)$$

**Modelo AR.** De la Tabla 3.3, se observa que la respectiva función de transferencia no tiene ceros, ( $q = 0$ ), y presenta solamente polos (todo polos), los cuales tienen respuesta a impulso con longitud infinita, estos pertenecen a la clase de los recursivos. La DEP de un proceso AR, haciendo  $q = 0$ , resulta en

$$S(\Omega) = \frac{\sigma^2}{|A(\Omega)|^2} = \frac{\sigma^2}{\mathbf{e}_p^H(\Omega) \mathbf{a} \mathbf{a}^H \mathbf{e}_p^H(\Omega)}$$

### 3.4.3. Estimación de los parámetros a partir de la función de correlación

Sea conocida la función de correlación propia de la sucesión aleatoria en análisis. Entonces, al multiplicar ambas partes de la ecuación (3.77a) por el factor  $x^*[n - m]$  y calculando la

esperanza matemática se obtiene

$$E\{x[n]x^*[n-k]\} = -\sum_{l=1}^p a_l E\{x[n-l]x^*[n-k]\} + \sum_{l=0}^q b_l E\{u[n-l]x^*[n-k]\}$$

que en términos de la función de correlación se escribe como

$$r_x[k] = -\sum_{l=1}^p a_l r_x[k-l] + \sum_{l=0}^q b_l r_{ux}[k-l]$$

la función de correlación mutua  $r_{ux}[k]$  entre las secuencias de entrada y salida se pueden expresar mediante los parámetros de la respuesta impulso  $h[m]$ :

$$\begin{aligned} r_{ux}[k] &= E\{u[n+k]x^*[n]\} = E\{u[n+k]u^*[n] + \sum_{m=1}^{\infty} h^*[m]u^*[n-m]\} \\ &= r_u[k] + \sum_{m=1}^{\infty} h^*[m]r_u[k+m] \end{aligned}$$

por cuanto  $u[n]$  es una sucesión de RBG con varianza  $\sigma^2$ , entonces,

$$r_{ux}[k] = \begin{cases} 0, & k > 0; \\ \sigma^2, & k = 0; \\ \sigma^2 h^*[-k], & k < 0. \end{cases}$$

De lo cual se obtiene la expresión que relaciona los coeficientes del modelo ARMA con la función de correlación propia de la señal discretizada aleatoria  $x[n]$

$$r_x[k] = \begin{cases} r_x[-k], & k < 0; \\ -\sum_{l=1}^p a_l r_x[k-l] + \sigma^2 \sum_{l=k}^q b_l h^*[l-k], & 0 \leq k \leq q; \\ -\sum_{l=1}^p a_l r_x[k-l], & k > q. \end{cases} \quad (3.80)$$

donde  $h[0] = 1$ , por definición, además con el fin de cumplir la condición de causalidad, el filtro debe tener respuesta a impulso  $h[k] = 0, \forall k < 0$ .

Cabe anotar que la relación en los parámetros del modelo ARMA y la función de correlación es no lineal, en particular, el sistema de ecuaciones (3.80) es no lineal, debido a la presencia del término  $\sum_{l=k}^q b_l h^*[l-k]$ . Sin embargo, cuando el modelo paramétrico es del tipo AR, el modelo tiene carácter lineal. Así, asumiendo en (3.80)  $b_l = \delta[l]$ , entonces se obtiene

$$r_x[k] = -\sum_{l=1}^p a_l r_x[k-l] + \sigma^2 h^*[-k], \quad 0 \leq k \leq q$$

Debido a que  $h^*[-k] = 0, \forall k > 0$ ,  $h^*[0] = \lim_{z \rightarrow \infty} H(z) = 1$ , entonces de (3.80) se tiene que

$$r_x[k] = \begin{cases} -\sum_{l=1}^p a_l r_x[k-l], & k > 0; \\ -\sum_{l=1}^p a_l r_x[-l] + \sigma^2, & k = 0; \\ r_x^*[-k], & k < 0. \end{cases} \quad (3.81)$$

Las expresiones (3.81) corresponden a las ecuaciones normales de Yule-Walker para un proceso AR [29]. Estas ecuaciones, en general, caracterizan la relación no lineal entre los parámetros del proceso AR y la función de correlación propia de la señal aleatoria analizada. Sin embargo, conocida la función de correlación propia se pueden determinar los parámetros del modelo AR al resolver un sistema de ecuaciones lineales lo cual se puede observar al plantear (3.81) en forma matricial:

$$\begin{bmatrix} r_x[0] & r_x[-1] & \cdots & r_x[-(p-1)] \\ r_x[1] & r_x[0] & \cdots & r_x[-(p-2)] \\ \vdots & \vdots & \ddots & \vdots \\ r_x[p-1] & r_x[p-2] & \cdots & r_x[0] \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_p \end{bmatrix} = \begin{bmatrix} r_x[1] \\ r_x[2] \\ \vdots \\ r_x[p] \end{bmatrix} \quad (3.82)$$

De esta manera, conocida la sucesión de valores  $r_x[k]$ ,  $-p \leq k \leq p$ , para la función de correlación propia, entonces los parámetros del modelo AR se pueden hallar como la solución del sistema de ecuaciones lineales (3.82). Cabe anotar que la matriz de función de correlación en (3.82) es Toeplitz y Hermitiana, debido a  $r_x[-k] = r_x^*[k]$ , por lo que la obtención de los valores  $\sigma^2$ ,  $a_1$ ,  $a_2, \dots, a_p$  se puede hacer empleando el algoritmo de *Levinson-Durbin* [29].

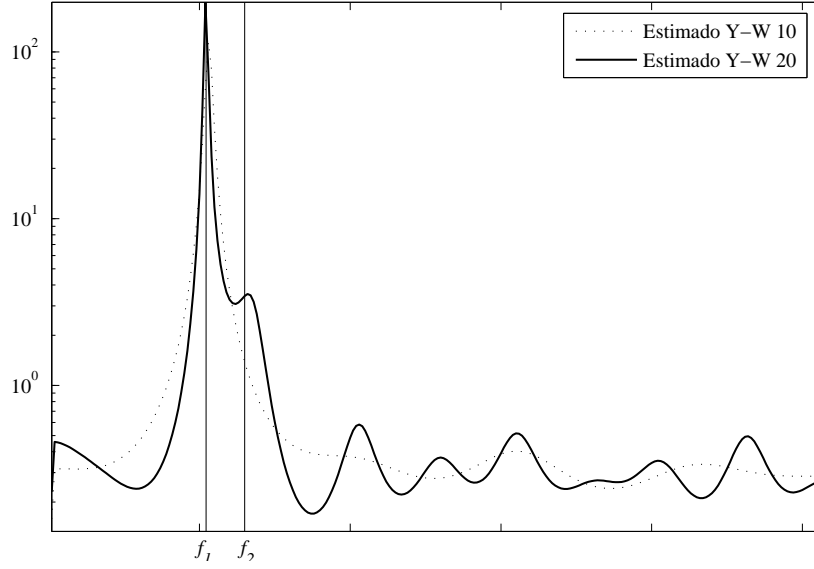
La relación de los valores, entre la función de correlación propia y los parámetros del modelo MA, se puede obtener de (3.80) asumiendo  $p = 0$  y teniendo en cuenta que para los sistemas no recursivos se cumple que  $h[k] = b_k, \forall 1 \leq k \leq q$ , entonces,

$$r_x[k] = \begin{cases} 0, & k > q; \\ \sigma^2 \sum_{l=k}^q b_l b_{l-k}^*, & 0 \leq k \leq q \\ 0, & k < 0 \end{cases} \quad (3.83)$$

Luego, la relación entre los parámetros MA y la función de correlación tiene carácter no lineal, condicionado por la convolución en (3.83). Otra forma conocida de estimar los parámetros del modelo AR, consiste en hallar directamente la solución del sistema de ecuaciones lineales, que se obtiene de la relación (3.80) para los  $p$  valores del índice  $k$ , ubicados en el intervalo  $q \leq k \leq q+p-1$ . Los valores de los parámetros AR se encuentran

de la solución del siguiente sistema de ecuaciones:

$$\begin{bmatrix} r_x[q] & r_x[q-1] & \cdots & r_x[q-p+1] \\ r_x[q+1] & r_x[q] & \cdots & r_x[q-p+2] \\ \vdots & \vdots & \ddots & \vdots \\ r_x[q+p-1] & r_x[q+p-2] & \cdots & r_x[q] \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_p \end{bmatrix} = \begin{bmatrix} r_x[1] \\ r_x[2] \\ \vdots \\ r_x[p] \end{bmatrix} \quad (3.84)$$



**Figura 3.10.** Ejemplo 3.10 calculado mediante el sistema modificado de Yule-Walker

La expresión (3.84) se denomina sistema de *ecuaciones normales de Yule-Walker* para los modelos ARMA o *sistema modificado de Yule-Walker*.

En la Figura 3.10, se muestra el ejemplo 3.10, calculado mediante el sistema modificado de Yule-Walker para diferentes órdenes del modelo.

## Problemas

**Problema 3.16.** Demostrar la igualdad de la ecuación (3.79).

**Problema 3.17.** Implementar el algoritmo de Levinson-Durbin para las ecuaciones normales de Yule-Walker para un proceso AR.

**Problema 3.18.** Determinar la media y la autocorrelación de la secuencia  $x[n]$ , que es salida de un proceso MA descrito por la ecuación en diferencias

$$x[n] = w[n] - 2w[n-1] + w[n-2]$$

donde  $w[n]$  es RBG con varianza  $\sigma_w^2$ .

**Problema 3.19.** Considerar un proceso MA descrito por la ecuación en diferencias

$$x[n] = w[n] + 0.81w[n-2]$$

donde  $w[n]$  es un RBG con varianza  $\sigma_w^2$ . Determinar los parámetros de los modelos AR de órdenes  $p = 2, 4, 8$ .

**Problema 3.20.** Hallar el correlograma y la DEP de la estimación del valor medio de longitud  $q$ :

$$x[n] = \frac{1}{q} (u[n] + u[n-1] \cdots + u[n-q+1])$$

**Problema 3.21.** Hallar la DEP, expresada mediante la respectiva función de correlación propia, para los procesos generados mediante un filtro formante con las siguientes funciones de transferencia

$$(a). H(z) = b_0 + b_1 z^{-1}. \quad (b). H(z) = b_0 + b_1 z^{-1} + b_2 z^{-2}.$$

**Problema 3.22.** Hallar el valor medio de la sucesión aleatoria descrita por la siguiente ecuación de diferencias:

$$x[n] = \alpha x[n-1] + u[n], \quad n \geq 0, \quad x[-1] = 0$$

donde  $u[n]$  es RBG discreto con media  $m_u$  y función de correlación  $r_u = \sigma_u^2 \delta[k]$

# Capítulo 4

## Procesos aleatorios de Markov

### 4.1. Definición y clasificación

Un proceso aleatorio  $\xi(t)$  se denomina *proceso de Markov* cuando para un valor fijo de  $\xi(u)$ , los demás valores aleatorios de  $\xi(t)$ ,  $t > u$ , no dependen de  $\xi(s)$ ,  $s < u$  [15]. Esto es, si para cualquier conjunto de argumentos de tiempo  $t_1, t_2, \dots, t_n$  definidos en el intervalo de análisis  $[0, T]$  la FDP condicional del último (extremo) valor  $\xi(t_n)$ , con  $\xi(t_1), \xi(t_2), \dots, \xi(t_{n-1})$ , depende solamente de  $\xi(t_{n-1})$ , con lo cual, para el conjunto  $\xi_1, \xi_2, \dots, \xi_n$ , es cierta la siguiente relación:

$$P(\xi(t_n) \leq \xi_n | \xi(t_1) = \xi_1, \dots, \xi(t_{n-1}) = \xi_{n-1}) = P(\xi(t_n) \leq \xi_n | \xi(t_{n-1}) = \xi_{n-1})$$

Una forma alterna de descripción del proceso de Markov está en su definición simétrica en los momentos de tiempo  $t_m, t_n, t_l$ :

$$\begin{aligned} P(\xi(t_m) \leq \xi_m, \xi(t_l) \leq \xi_l | \xi(t_n) = \xi_n) \\ = P(\xi(t_m) \leq \xi_m | \xi(t_n) = \xi_n) P(\xi(t_l) \leq \xi_l | \xi(t_n) = \xi_n) \end{aligned} \quad (4.1)$$

lo cual implica, que para un estado determinado del proceso, en el momento actual de tiempo  $t_n$ , los estados futuro  $t_l$  y pasado  $t_m$  cumplen la condición de independencia estadística.

De las anteriores definiciones, para los procesos de Markov, se deduce que la FDP con dimensión  $n$  que brinde la mayor información se puede representar en la forma,

$$p(\xi_1, \xi_2, \dots, \xi_n) = p(\xi_1) \prod_{i=1}^{n-1} p(\xi_{i+1} | \xi_i) \quad (4.2)$$

esto es, cualquier FDP con dimensión  $n$  de un proceso de Markov, se puede hallar empleando (4.2), cuando se conocen la FDP con dimensión simple del proceso y las densidades de probabilidad (o probabilidades) condicionales de cambio o *transición*. En forma general, la evolución de las probabilidades de cambio,  $P(\xi(t) \leq \xi | \xi(t_0) = \xi_0)$ , se describe por la

ecuación:

$$\frac{d}{dt}P = \mathcal{K}\{P\} \quad (4.3)$$

donde  $\mathcal{K}$  es un cierto operador lineal (matriz, diferenciador, etc.). La descripción operacional del sistema en (4.3), permite analizar el comportamiento de los procesos de Markov, empleando métodos conocidos de solución para las respectivas ecuaciones diferenciales, en las cuales el carácter de las tareas físicas desarrolladas pueden ser variadas, de acuerdo a las condiciones dadas iniciales y límite del sistema.

Sea el comportamiento de un sistema que se describe por alguna ecuación que cumpla (4.1), para unas condiciones iniciales dadas (el estado del sistema en el momento inicial  $t_0$ ). Si no se asume alguna otra restricción se debe hallar directamente la solución de (4.3), en caso contrario las condiciones límite deben reflejar las restricciones de solución. Las mismas restricciones tienen diferente naturaleza. Por ejemplo, un punto que refleja el comportamiento del sistema, que comienza su movimiento desde el estado inicial del sistema  $\xi(t_0) = \xi_0$ , cuando llegue al límite  $\xi(t) = c$  puede absorberse y el funcionamiento del sistema se detiene (Figura 4.1, trayectoria 1).

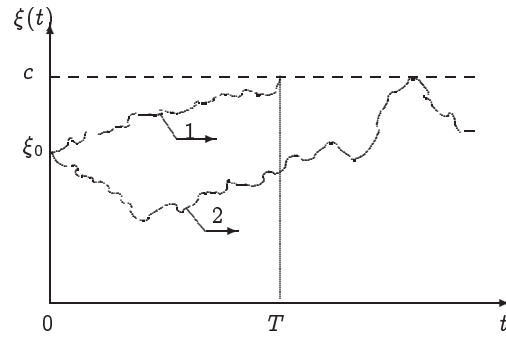
En otro caso, el punto puede reflejarse de la frontera  $c$  (trayectoria 2), y así sucesivamente. Si la frontera  $c$  es absorbente, entonces además de estimar la probabilidad de transición  $P$  para  $\xi < c$ , se puede hallar la probabilidad de absorción en algún tiempo  $T$ , o en la esperanza matemática, varianza o algún otro momento de tiempo  $T$ , en el cual el sistema por primera vez alcanza la frontera  $c$ .

La ecuación (4.3) tiene las siguientes formas de caracterización de los procesos de Markov:

1. Procesos de tiempo discreto y espacio discreto de fase,
2. Procesos de tiempo discreto y espacio continuo de fase
3. Procesos de tiempo continuo y espacio discreto de fase,
4. Procesos de tiempo continuo y espacio continuo de fase.

#### 4.1.1. Cadenas de Markov

Sea una señal aleatoria  $\theta(t)$  que toma una cantidad finita  $K$  de valores discretos, pertenecientes al conjunto  $\{\vartheta_1, \dots, \vartheta_K\}$ , tal que en los momentos determinados del tiempo ( $t_0 < t_1 < t_2 < \dots$ ) los valores del proceso aleatorio cambian de manera abrupta (con probabilidades de



**Figura 4.1.** Influencia de la frontera en la trayectoria del proceso



cambio conocidas), esto es, los cambios forman la sucesión aleatoria  $\theta_0 \rightarrow \theta_1 \rightarrow \theta_2, \dots$ , siendo  $\theta_n = \theta(t_n)$  el valor de la sucesión para el intervalo  $n$  de tiempo.

Una *cadena compuesta de Markov* de orden  $m \in \mathbb{Z}$  implica que la probabilidad de un nuevo valor del proceso depende, sola y estrictamente, de los  $m$  valores que lo anteceden:

$$P(\theta_n | \theta_0, \theta_1, \dots, \theta_{n-1}) = P(\theta_n | \theta_{n-m}, \dots, \theta_{n-1})$$

En general, una cadena de orden  $m$  se puede representar por una cadena simple ( $m = 1$ ), razón por la cual, de forma usual, el análisis de los procesos de Markov se realiza sobre cadenas simples, para las cuales las probabilidades conjuntas y finitas, determinadas a partir de (4.2), se expresan por medio de las *probabilidades de cambio*  $P(\theta_\mu | \theta_{\mu-1})$ :

$$P(\theta_0, \theta_1, \dots, \theta_n) = P(\theta_0) \prod_{\mu=1}^n P(\theta_\mu | \theta_{\mu-1})$$

Las cadenas simples de Markov, en la mayoría de los casos prácticos, se usan para la estimación de los diferentes valores de probabilidad en los momentos del tiempo  $t_n > t_0$ , y particularmente, cuando  $n \rightarrow \infty$ . En la descripción de las cadenas, es usual el empleo de las siguientes notaciones para los vectores columna de la probabilidades marginales y matrices de las probabilidades condicionales (ambas probabilidades deben cumplir con las condiciones de realización física y normalización):

$$\left\{ \begin{array}{ll} \mathbf{P}(n) &= \{p_k(n) = P(\theta_n = \vartheta_k) : k = 1, \dots, K\}, \quad n = 0, \dots, N \\ \mathbf{\Pi}(\mu, n) &= \{\pi_{mk}(\mu, n) = P(\theta_n = \vartheta_k | \theta_\mu = \vartheta_m) : m, k = 1, \dots, K\}, \quad 0 \leq \mu \leq n, n = 0, \\ p_k(n) &\geq 0, \quad \sum_{k=1}^K p_k(n) = 1, \quad n = 0, \dots, N \\ \pi_{mk}(\mu, n) &\geq 0, \quad \sum_{k=1}^K \pi_{mk}(\mu, n) = 1, \quad m = 1, \dots, K \end{array} \right. \quad (4.4)$$

El valor de  $p_k(n)$  es el valor de la probabilidad marginal de  $\vartheta_k$  en el instante  $n$ , cuando  $t = t_n$ . Mientras,  $\pi_{mk}(\mu, n)$  determina la probabilidad condicional del valor  $\vartheta_k$  en  $t_n$ , si en el momento anterior  $t_\mu < t_n$  el valor del proceso fue igual a  $\vartheta_m$ . Basados en la relación de probabilidad completa (2.3), para el sistema de (4.4), se escribe la ecuación de Markov:

$$\begin{aligned} \mathbf{P}(n) &= \mathbf{\Pi}^\top(\mu, n) \mathbf{P}(\mu) \\ \mathbf{\Pi}(\mu, n) &= \mathbf{\Pi}(\mu, m) \mathbf{\Pi}(m, n) = \prod_{i=0}^{n-\mu-1} \mathbf{\Pi}(\mu + i, \mu + i + 1), \quad n > m > \mu \geq 0 \end{aligned} \quad (4.5)$$

con lo cual, para la determinación de la matriz  $\mathbf{\Pi}(\mu, n)$ ,  $\mu \leq n$ , es suficiente conocer la sucesión de matrices por cada intervalo de las probabilidades de cambio. Más aún, se puede afirmar que la descripción completa de una cadena simple de Markov se alcanza con la probabilidad del estado inicial y la matriz de las probabilidades de transición de la sucesión:

$$\mathbf{\Pi} = \begin{bmatrix} \pi_{11} & \pi_{12} & \cdots & \pi_{1K} \\ \pi_{21} & \pi_{22} & \cdots & \pi_{2K} \\ \cdots & \cdots & \cdots & \cdots \\ \pi_{K1} & \pi_{K2} & \cdots & \pi_{KK} \end{bmatrix} \quad (4.6)$$

En las cadenas *homogéneas*, la probabilidad de cambio es invariable al tiempo y depende solamente de la diferencia de argumentos:

$$\mathbf{\Pi}(\mu, n) = \mathbf{\Pi}(n - \mu), \quad n > \mu \geq 0$$

A partir de (4.5) y al notar  $\mathbf{\Pi} = \mathbf{\Pi}(1)$ , se obtiene que la matriz de transición en una cadena simple y homogénea de Markov, después de  $n$  pasos, es igual a:

$$\mathbf{\Pi}(n) = \mathbf{\Pi}^n, \quad \mathbf{P}^\top(n) = \mathbf{P}^\top(0) \mathbf{\Pi}^n \quad (4.7)$$

**Ejemplo 4.1.** Hallar la probabilidad de que después de  $n$  pasos la variable aleatoria pasa del estado  $\vartheta_i$  al estado  $\vartheta_j$ , si el proceso corresponde a una sucesión simple homogénea de Markov con  $K$  estados.

La matriz de transición corresponde a la expresión (4.6), de la cual se observa que la probabilidad de que la variable aleatoria pase del estado  $\vartheta_i \rightarrow \vartheta_j$  en un solo paso es igual a  $\pi_{ij}$ . Al punto  $j$  en dos pasos, la sucesión aleatoria puede llegar de dos maneras:  $i \rightarrow l; l \rightarrow j$ . La probabilidad de que la sucesión tome exactamente esa trayectoria es igual a  $\pi_{il}\pi_{lj}$ . Sin embargo, la cantidad de posibles puntos es  $K$ , por lo que teniendo en cuenta el teorema (2.1), se obtiene

$$\pi_{ij}(2) = \sum_{l=1}^K \pi_{il}\pi_{lj}, \quad \mathbf{\Pi}(2) = \{\pi_{ij}(2) : i, j = 1, \dots, K\}$$

La sumatoria abierta del elemento  $\pi_{ij}(2)$  es igual a,

$$\pi_{ij}(2) = \sum_{l=1}^K \pi_{il}\pi_{lj} = \pi_{i1}\pi_{1j} + \pi_{i2}\pi_{2j} + \cdots + \pi_{iK}\pi_{Kj}$$

la cual corresponde a la expresión para la multiplicación de matrices, esto es,

$$\mathbf{\Pi}(2) = \mathbf{\Pi}(1) \mathbf{\Pi}(1) = \mathbf{\Pi}^2$$

De manera similar, se obtiene la matriz para  $3, 4, \dots, n$  pasos:  $\mathbf{\Pi}(n) = \mathbf{\Pi}^n$ .

**Ejemplo 4.2.** Sea la matriz de transición

$$\mathbf{\Pi} = \begin{bmatrix} 1/2 & 1/2 \\ 2/5 & 3/5 \end{bmatrix}$$

Hallar la matriz de transición para dos pasos y calcular las probabilidades de que el sistema llega al estado 1 en un paso, además al estado 2 en dos pasos, si se conoce el estado inicial:  $p_0(1) = 1$  y  $p_0(2) = 0$ .

Por cuanto,  $\mathbf{\Pi}(2) = \mathbf{\Pi}^2$ , entonces

$$\mathbf{\Pi}(2) = \begin{bmatrix} 1/2 & 1/2 \\ 2/5 & 3/5 \end{bmatrix} \begin{bmatrix} 1/2 & 1/2 \\ 2/5 & 3/5 \end{bmatrix} = \begin{bmatrix} \pi_{11}(2) & \pi_{12}(2) \\ \pi_{21}(2) & \pi_{22}(2) \end{bmatrix} = \begin{bmatrix} 9/20 & 11/20 \\ 11/25 & 14/25 \end{bmatrix}$$

las probabilidades absolutas  $p_j(n)$ , teniendo en cuenta el vector dado inicial  $\mathbf{P}^T(0) = [1 \ 0]$ , se hallan de la expresión (4.7),

$$\begin{aligned} \mathbf{P}^T(1) &= [1 \ 0] \begin{bmatrix} 1/2 & 1/2 \\ 2/5 & 3/5 \end{bmatrix} = [1/2 \ 1/2] \\ \mathbf{P}^T(2) &= [1 \ 0] \begin{bmatrix} 9/20 & 11/20 \\ 11/25 & 14/25 \end{bmatrix} = [9/20 \ 11/20] \end{aligned}$$

La lectura de los resultados obtenidos es la siguiente: si la sucesión inicialmente se encontraba en el primer estado, entonces la probabilidad de que en un paso la trayectoria se quede en este mismo valor de estado es igual a  $1/2$ , e igual a  $1/2$ , si toma el segundo estado. Para estas mismas condiciones iniciales, la probabilidad de que en dos pasos la sucesión tome el segundo estado es igual a  $9/20$ , mientras la probabilidad de tomar el segundo estado es  $9/20$ .

Una cadena homogénea se denomina *estacionaria* si se cumple que  $\mathbf{P}(n) = \mathbf{P}$ , lo cual implica que no depende del valor normalizado del tiempo  $n$ . En forma general, las probabilidades *finales*  $\mathbf{P}$ , si existen, se hallan como resultado del límite de la matriz de transición,

$$\mathbf{P} = \lim_{n \rightarrow \infty} \mathbf{P}(n) \quad (4.8)$$

Sin embargo, si se tiene que las probabilidades iniciales  $\mathbf{P}(0)$  coinciden con las probabilidades finales correspondientes  $\mathbf{P}$ , entonces la cadena de Markov es estacionaria a partir del momento de tiempo  $t_0$ .

Las probabilidades finales deben satisfacer el siguiente sistema lineal de ecuaciones algebraicas de orden  $K$ :

$$(\mathbf{I} - \mathbf{\Pi}^T) \mathbf{P} = \mathbf{0} \quad (4.9)$$

además de la condición adicional  $\sum_{k=1}^K p_k = 1, p_k \geq 0$ , siendo  $\mathbf{I}$  la matriz unitaria. Debido a la condición (4.4) las  $K$  ecuaciones de (4.9) son linealmente dependientes. Por lo tanto, las  $K$  probabilidades finales se deben determinar a partir de las ecuaciones de (4.9).

La clasificación de los estados de un proceso de Markov depende de si, dado un estado, el proceso puede llegar hasta cualquier otro estado. Un estado  $\vartheta_j$  se denomina *sin retorno*

si existe un estado  $\mathfrak{v}_k, k \neq i$ , después de una cantidad de pasos  $n$ , tales que  $\pi_{jk}(n) > 0$ , pero  $\pi_{kj} = 0, \forall m$ . Los estados *con retorno* suponen el retorno desde cualquier estado del proceso, inclusive si la cantidad de pasos en la dirección de ida y en la de retorno no son iguales. De otra parte, la pareja de estados se denomina *conectiva*, si dados  $n, m$  se cumplen las condiciones  $\pi_{jk}(n) > 0$  y  $\pi_{kj}(m) > 0$ . El conjunto de estados con retorno y conectivos se denomina *ergódico*. Por cierto, si dado un valor suficientemente grande de pasos  $n_0$ , se cumple que todos los elementos de la matriz  $\mathbf{\Pi}^n$  son todos positivos para cualquier  $n > n_0$ , entonces el proceso se define como *regularmente ergódico*.

En los procesos ergódicos, se define la matriz fundamental, por medio de la cual se pueden hallar otras características estadísticas básicas del proceso, en la forma

$$\mathbf{Z} = (\mathbf{I} + \mathbf{\Pi} + \mathbf{G})^{-1}$$

donde  $\mathbf{G}$  es la matriz, en la cual cada fila es igual al vector  $\mathbf{P}^T$  de las probabilidades finales en (4.8).

**Ejemplo 4.3.** Hallar el valor asintótico de las probabilidades absolutas  $\mathbf{P}(n)$ , para  $n \rightarrow \infty$ , de una cadena ergódica de Markov.

La ecuación (4.7) determina las probabilidades absolutas después de  $n$  pasos de la sucesión, para la cual se analiza la siguiente relación límite por cada uno de sus términos elementales,

$$\lim_{n \rightarrow \infty} p_j(n) = \lim_{n \rightarrow \infty} \sum_{i=1}^K p_i(0) \pi_{ij}(n) = \sum_{i=1}^K p_i(0) \lim_{n \rightarrow \infty} \pi_{ij}(n)$$

Al tener en cuenta, los valores finales de probabilidad, en correspondencia con la definición de ergodicidad, se obtiene,

$$\lim_{n \rightarrow \infty} p_j(n) = \sum_{i=1}^K p_i(0) \pi_{ij} = \pi_{ij} \sum_{i=1}^K p_i(0)$$

que muestra la validez de (4.8), debido a que  $\sum_{i=1}^K p_i(0) = 1$ .

**Ejemplo 4.4.** Establecer si la matriz de transición del ejemplo 4.2 es o no ergódica; en tal caso, calcular las probabilidades finales.

En el ejemplo 4.2 se obtuvo la matriz de transición para 2 pasos igual a  $\mathbf{\Pi}(2) = \mathbf{\Pi}^2$ , con todos su elementos positivos,  $\{\pi_{ij}(2) \geq 0\}$ . De lo cual se puede afirmar que los elementos de la matriz  $\mathbf{\Pi}^n$  para  $n \rightarrow \infty$ , también serán estrictamente positivos y, por lo tanto, la cadena es ergódica.

En cuanto a las probabilidades finales, éstas se calculan resolviendo las condiciones en (4.9), que para cada uno de los elementos toma la forma,

$$\begin{cases} p_i &= \sum_{l=1}^K p_l \pi_{li} \\ \sum_{i=1}^K p_i &= 1, p_i > 0 \end{cases}$$

En el caso concreto, se tiene las siguientes ecuaciones:

$$\begin{cases} p_1 = p_1 \pi_{11} + p_2 \pi_{21} = p_1 1/2 + p_2 2/5 \\ p_2 = p_1 \pi_{12} + p_2 \pi_{22} = p_1 1/2 + p_2 3/5 \\ p_1 + p_2 = 1 \end{cases}$$

De la primera ecuación se tiene  $p_1 1/2 = p_2 2/5$ ,  $p_1 1/2 = p_2 2/5$ ,  $p_1 = 1 - p_2$ , entonces,  $p_1 = 4/9$ ,  $p_2 = 5/9$ . Los resultados obtenidos al resolver el problema 4.1, en el primer caso, muestran que los elementos de las matrices  $\Pi^3, \Pi^4, \dots$ , convergen a los resultados obtenidos.

#### 4.1.2. Procesos discretos de Markov

Sea un proceso aleatorio  $\theta(t)$  que toma solamente valores discretos  $\{\vartheta_k : k = 1, \dots, K\}$ , pero cuyo cambio no ocurre en momentos fijos, sino en valores aleatorios del tiempo (Figura 4.2), caso en el cual las notaciones de (4.4) se formulan de nuevo así:

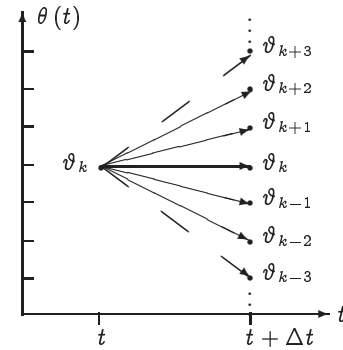
$$\begin{cases} \mathbf{P}(t) &= \{p_k(t) = P(\theta(t) = \vartheta_k) : k = 1, \dots, K\} \\ \Pi(t_0, t) &= \{\pi_{mk}(t_0, t) = P(\theta(t) = \vartheta_k \mid \theta(t_0) = \vartheta_m) : m, k = 1, \dots, K\}, \quad 0 \leq t_0 \leq t \\ p_k(t) &\geq 0, \quad \sum_{k=1}^K p_k(t) = 1 \\ \pi_{mk}(t_0, t) &\geq 0, \quad \sum_{k=1}^K \pi_{mk}(t_0, t) = 1, \quad j = 1, \dots, K \end{cases} \quad (4.10)$$

Similar a (4.5), para las probabilidades de cambio se cumple la *ecuación de Chapman-Kolmogorov*:

$$\Pi(t_0, t + \Delta t) = \Pi(t_0, t) \Pi(t, t + \Delta t), \quad t > t_0, \quad \Delta t > 0 \quad (4.11)$$

En los procesos discretos de Markov se asume la propiedad de *ordinariedad* [15], según la cual, para intervalos relativamente pequeños de tiempo  $\Delta t$  entre los discretos, se considera que la probabilidad  $\pi_{kk}$  de que el valor actual no cambie, excede la probabilidad de transición de este valor, esto es,

$$\begin{aligned} \pi_{kk}(t, t + \Delta t) &= 1 + a_{kk}(t) \Delta t + o(\Delta t) \\ \pi_{mk}(t, t + \Delta t) &= a_{kl}(t) \Delta t + o(\Delta t) \end{aligned} \quad (4.12)$$



**Figura 4.2.** Procesos discretos

donde  $o(\Delta t)$  denota los términos residuales, considerados insignificantes, mayores al de primer orden con relación a  $\Delta t$ :  $\lim_{\Delta t \rightarrow 0} \frac{o(\Delta t)}{\Delta t} = 0$ .

De acuerdo con la condición de estacionariedad (4.7), de la última expresión se tiene que

$$a_{kk}(t) = - \sum_{m \neq k} a_{km}(t) \leq 0, \quad a_{km}(t) \geq 0 \quad (4.13)$$

Además, tienen lugar las siguientes igualdades:

$$\Pi(t_0, t_0) = \mathbf{I}, \quad \Pi(t_0, t + \Delta t) = \mathbf{I} + \mathbf{A}(t) \Delta t + \mathbf{o}(\Delta t) \quad (4.14)$$

donde  $\mathbf{A}(t)$  corresponde a la matriz *infinitesimal de probabilidades*. Reemplazando la igualdad (4.14) en (4.11), y haciendo  $\Delta t \rightarrow 0$  se obtiene la *ecuación de Kolmogorov en el sentido directo* (esto es, orientada en el sentido natural de desarrollo del tiempo):

$$\frac{\partial}{\partial t} \Pi(t_0, t) = \Pi(t_0, t) \mathbf{A}(t) \quad (4.15)$$

cuya solución general, con condiciones iniciales (4.14) y cuando  $\mathbf{A}(t) = \mathbf{A} = \text{const.}$ , corresponde a la matriz exponencial

$$\Pi(t_0, t) = \exp(\mathbf{A}(t - t_0)) \quad (4.16)$$

Un proceso discreto de Markov conserva su tipo, cuando se analiza su desarrollo en el sentido contrario del tiempo, para el que se tiene la *ecuación inversa de Kolmogorov*:

$$\frac{\partial}{\partial t_0} \Pi(t_0, t) = -\mathbf{A}(t_0) \Pi(t_0, t), \quad t \geq t_0$$

que tiene como solución, asumiendo  $\mathbf{A}(t) = \mathbf{A} = \text{const.}$ , la matriz de forma exponencial

$$\mathbf{P}(t_0, t) = \exp(\mathbf{A}^\top(t - t_0)) \mathbf{P}(t_0)$$

A su vez, un proceso discreto de Markov se denomina *homogéneo* si la matriz de transición depende solamente de la diferencia  $\tau = t - t_0$ , entonces:

$$\Pi(t_0, t) = \Pi(t - t_0) = \Pi(\tau)$$

De la condición (4.12), se observa que para un proceso discreto de Markov homogéneo, la matriz de transición de probabilidades infinitesimales  $\mathbf{A}(t) = \mathbf{A}$  no depende del tiempo y el sistema de ecuaciones diferenciales de (4.15) se simplifica hasta la expresión:

$$\frac{\partial}{\partial \tau} \Pi(\tau) = \Pi(\tau) \mathbf{A} \quad (4.17)$$

Un proceso discreto de Markov se considera *ergódico*, si para  $\tau \rightarrow \infty$  existen los valores límites de las probabilidades de estados  $\mathbf{P} = \lim_{\tau \rightarrow \infty} \mathbf{P}(\tau)$ . Así mismo, dada la condición  $\sum_{k=1}^K p_k = 1$ , las probabilidades de los valores estacionarios se determinan del sistema algebraico de ecuaciones con  $K - 1$  variables:  $\mathbf{A}^\top \mathbf{P} = \mathbf{0}$ .

**Ejemplo 4.5.** Hallar la probabilidad  $P_k(t)$  de que un proceso discreto homogéneo de Markov en el momento de tiempo  $t$  se encuentre en el estado  $\vartheta_k$ ,  $k = 1, \dots, K$ .

La solución básicamente consiste en hallar la probabilidad  $P_k(t + \Delta t)$  de que el proceso tome el valor  $\vartheta_k$  en un tiempo  $(t + \Delta t)$ , para luego hallar la relación límite de cambio. En este caso, se considera que el proceso puede llegar  $\vartheta_k$ , con probabilidad de cambio  $\pi_{lk}(t, t + \Delta t)$  a partir del estado  $\vartheta_l$  en los siguientes casos:

1.  $\vartheta_k \rightarrow \vartheta_k$ , el proceso no cambia de estado con probabilidad de transición  $\pi_{kk}(t, t + \Delta t)$ .
2.  $\vartheta_{k-1} \rightarrow \vartheta_k \rightarrow \pi_{k-1,k}(t, t + \Delta t)$
3.  $\vartheta_{k+1} \rightarrow \vartheta_k \rightarrow \pi_{k+1,k}(t, t + \Delta t)$
4.  $\vartheta_{k \pm q} \rightarrow \vartheta_k \rightarrow \pi_{k \pm q,k}(t, t + \Delta t)$ , para  $q \geq 2$ , cuando el sistema varia, bien a niveles superiores o bien sea a niveles inferiores, en dos o más estados.

Las respectivas probabilidades de transición se calculan de la ecuación (4.12) que, teniendo en cuenta la homogeneidad del proceso, toma la forma

$$\begin{cases} \pi_{k,k+1}(\Delta t) &= \alpha_k \Delta t + o(\Delta t), \alpha \geq 0 \\ \pi_{k,k-1}(\Delta t) &= \beta_k \Delta t + o(\Delta t), \beta \geq 0, \alpha\beta > 0 \\ \pi_{k,\pm q}(\Delta t) &= o(\Delta t), q \geq 2 \end{cases}$$

con lo cual se tiene,

$$\begin{aligned} \pi_{k,k}(\Delta t) &= P_k(t)(1 - \pi_{k,k+1}(\Delta t) - \pi_{k,k-1}(\Delta t) - \pi_{k,\pm q}(\Delta t)) \\ &= P_k(t)(1 - \alpha_k \Delta t - \beta_k \Delta t - 3o(\Delta t)) \\ \pi_{k,k-1}(\Delta t) &= P_{k-1}(t)(\beta_{k-1} \Delta t + o(\Delta t)) \\ \pi_{k,k+1}(\Delta t) &= P_{k+1}(t)(\alpha_{k+1} \Delta t + o(\Delta t)) \\ \pi_{k,k \pm q}(\Delta t) &= P_{k \pm q}(t)o(\Delta t) \end{aligned}$$

Por cuanto, los cuatro eventos son independientes y no excluyentes, entonces, la probabilidad  $P_k(t + \Delta t)$  es igual a la suma de las anteriores cuatro probabilidades de transición,

$$P_k(t + \Delta t) = P_k(t) - (\alpha_k + \beta_k) P_k(t) \Delta t + \alpha_{k-1} P_{k-1}(t) \Delta t + \beta_{k+1} P_{k+1}(t) \Delta t + O_k(\Delta t) \quad (1)$$

donde  $O_k(\Delta t) = -3P_k(t)o(\Delta t) + P_{k-1}(t)o(\Delta t) + P_{k+1}(t)o(\Delta t) + P_{k \pm q}(t)o(\Delta t)$  es el término relacionado con la probabilidad de transición del estado actual a otro estado que se diferencie en dos o más niveles, y la cual se considera infinitamente pequeña, esto es, se asume el cambio suficientemente lento del proceso, tal que no se consideran saltos de más de un nivel de diferencia.

De la ecuación (1), asumiendo  $O_k(\Delta t) = 0$ , entonces se tiene que

$$\frac{P_k(t + \Delta t) - P_k(t)}{\Delta t} = -(\alpha_k + \beta_k) P_k(t) + \alpha_{k-1} P_{k-1}(t) + \beta_{k+1} P_{k+1}(t)$$

que al realizar el cambio límite da como resultado la expresión,

$$\frac{dP_n(t)}{dt} = -(\alpha_k + \beta_k) P_k(t) + \alpha_{k-1} P_{k-1}(t) + \beta_{k+1} P_{k+1}(t) \quad (2)$$

De forma similar se obtiene la relación para probabilidad del estado más bajo del sistema  $\vartheta_1$ , asumiendo  $P_{k-1}(t) = 0$  y  $\beta_1 = 0$ , mientras para el estado más alto,  $\vartheta_K$ , se asume,  $P_{k+1}(t) = 0$  y  $\alpha_K = 0$ . La generalización de (2) corresponde a la ecuación (4.17) para un proceso discreto homogéneo con  $K$  estados.

Cabe anotar que la forma de cambio de los parámetros  $\alpha_k = \alpha(k)$ ,  $\beta_k = \beta(k)$  determina la dependencia en el tiempo de  $P_k(t)$  (o  $P_k$  en los procesos ergódicos). Así por ejemplo, si

se tiene  $\beta_k = 0$  y  $\alpha_k = \lambda$ , entonces la probabilidad  $P_k(t)$  corresponde a la FDP de Poisson con parámetro  $\lambda$ , mostrada en el ejemplo 2.8. En cambio, cuando  $\alpha_k = \alpha$ , y  $\beta_k = k\beta$ , asumiendo la ergodicidad del proceso, la probabilidad  $P_k(t)$  corresponde a la FDP de Erlang, que corresponde a un caso particular de la  $\Gamma$ -densidad (2.28).

Cuando uno de los parámetros correspondientes a las probabilidades de transición, presentes en la ecuación (1), se asume igual a cero, implica que el proceso tiene una sola dirección de desarrollo: si  $\alpha(k) = 0, \forall k$ , entonces, el proceso nunca crece y su desarrollo corresponde a la disminución del orden del estado (desintegración o muerte), cuando  $\beta(k) = 0, \forall k$ , el proceso nunca decrece y aumenta de forma progresiva el orden de los estados (nacimiento).

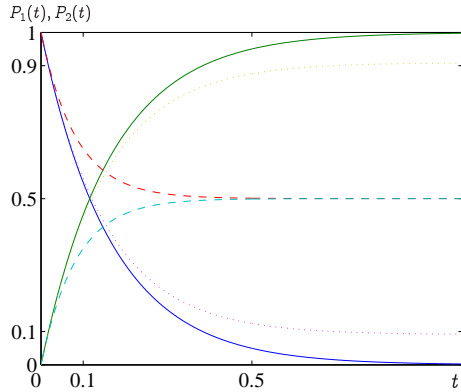
**Ejemplo 4.6.** Un proceso discreto homogéneo de Markov puede encontrarse en uno de dos posibles estados:  $\{\vartheta_k : k = 1, 2\}$ . El cambio del primer estado  $\vartheta_1 \rightarrow \vartheta_2$  ocurre con probabilidad  $\alpha dt$ , mientras el cambio  $\vartheta_2 \rightarrow \vartheta_1$  ocurre con probabilidad  $\beta dt$ . Hallar las respectivas probabilidades  $P_k(t)$  de que el proceso se encuentre en cada uno de los dos estados, asumiendo que el proceso en el tiempo  $t = 0$  se encuentra en el estado  $\vartheta_1$ .

De las condiciones del proceso se tiene que  $A_{12} = \alpha$  y  $A_{21} = \beta$  y, teniendo en cuenta (4.13),  $A_{11} = -\alpha$ ,  $A_{22} = -\beta$ . Por lo tanto, la ecuación (4.15) para las probabilidades de estado definidas en (4.10)  $p_1(t) = P_1(t)$  y  $p_2(t) = P_2(t)$ , toma la forma

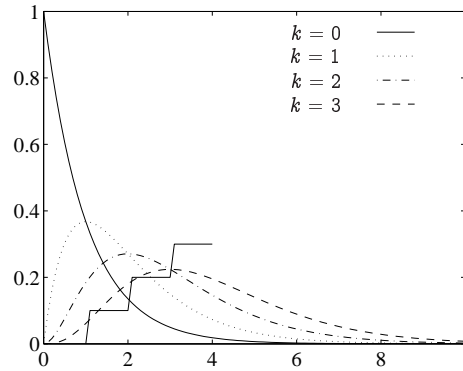
$$dP_1/dt = -\alpha P_1 + \beta P_2, \quad dP_2/dt = \alpha P_1 - \beta P_2 \quad (3)$$

Por cuanto,  $P_1 + P_2 = 1$ , se emplea una sola ecuación en (3), por ejemplo, la primera,

$$dP_1/dt = -\lambda P_1 + \beta, \quad \lambda = \alpha + \beta$$



(a) Proceso con dos estados



(b) Proceso de nacimiento y varios estados

**Figura 4.3.** Dependencia de las probabilidades de estado en el tiempo

Inicialmente, se asume el valor de  $P_1 = 1$ , para  $t = 0$ , esto es, al principio el proceso con certeza se encuentra en el estado  $\vartheta_1$ . Entonces, de (4.16) se obtiene la solución

$$P_1(t) = e^{-\lambda t} + \frac{\beta}{\lambda} (1 - e^{-\lambda t}), \quad P_2(t) = \frac{\alpha}{\lambda} (1 - e^{-\lambda t})$$

El análisis de ambas funciones de probabilidad, cuando  $t \rightarrow \infty$ , muestra que se obtienen los valores asintóticos constantes, que no dependen de las condiciones iniciales,

$$P_1 = \beta/\lambda, \quad P_2 = \alpha/\lambda$$



de lo cual se establece que el proceso es ergódico.

Como antes se dijo, los coeficientes  $\alpha$  y  $\beta$  determinan la forma de las probabilidades de estado del proceso, como se observa en la Figura 4.3(a), en la cual se muestran ejemplos de cálculo para  $\alpha = \beta$  (línea  $--\rightarrow$ ),  $\alpha \gg \beta$  ( $\cdots$ ) y  $\beta = 0$  (con línea continua). Este último caso, que corresponde a un proceso de nacimiento,  $\vartheta_k \rightarrow \vartheta_{k+1}$ , se puede generalizar para  $\alpha(k) = \alpha$  y  $k$  estados,

$$P_k(t) = \frac{(\alpha t)^k}{k!} e^{-\alpha t}$$

Las respectivas dependencias para  $k = 0, 1, 2, 3$  se muestran en la Figura 4.3(b), en la cual además se muestra el estado más probable del procesos en cada momento del tiempo.

#### 4.1.3. Procesos continuos de Markov

El dominio del proceso aleatorio con valores continuos  $x(t)$  y el dominio de su definición  $[0, T]$  son conjuntos continuos, lo cual no significa que el proceso de Markov sea continuo, más bien, el término proceso de valores continuos significa que sus valores, definidos en un dominio continuo, cambian de manera abrupta en forma incremental en un dominio continuo de tiempo.

Sea la FDP  $p_\xi(x(t_k)|x(t_1), \dots, x(t_{k-1}))$  para la señal aleatoria  $\xi(t)$ , asumiendo que son conocidos todos los valores del proceso correspondientes a los momentos previos,  $t_1, \dots, t_{k-1}$ . La suposición Markoviana (4.2), en el caso continuo, se escribe en la forma,

$$p_\xi(x(t_k)|x(t_1), \dots, x(t_{k-1})) = p_\xi(x(t_k)|x(t_{k-1}))$$

En este caso, la FDP conjunta de los  $k$  valores considerados del proceso es igual a,

$$\begin{aligned} p_\xi(x(t_1), \dots, x(t_k)) &= p_\xi(x(t_k)|x(t_1), \dots, x(t_{k-1})) p_\xi(x(t_1), \dots, x(t_{k-1})) = \\ &= p_\xi(x(t_1)|x(t_{k-1})) p_\xi(x(t_{k-1})|x(t_{k-2})) \cdots \\ &\cdots p_\xi(x(t_2)|x(t_2)) p_\xi(x(t_1)) \end{aligned}$$

esto es, la FDP de los valores de un proceso de Markov de cualquier dimensión, se puede desdoblar a través de densidades de probabilidad de dos dimensiones del tipo  $p_\xi(x(t_n)|x(t_{n-1}))$  y la FDP  $p_\xi(t_1)$ , correspondiente al inicio del proceso.

De otra parte, teniendo en cuenta la relación unívoca entre las FDP condicionales y marginales del proceso, dadas por el teorema de Bayes 2.3, entonces se puede afirmar que todas las funciones de probabilidad de un proceso de Markov, de forma unívoca, también pueden expresarse mediante una FDP con dos dimensiones, correspondientes a dos valores cualesquiera del tiempo. A su vez, esta FDP de dos dimensiones caracteriza de manera completa el proceso.

Debido a la dinámica propia del proceso de Markov en el tiempo, es importante determinar la probabilidad de que la señal aleatoria  $\xi(t)$  con valor  $x_1$  para el momento de tiempo  $t_1$ , en el momento de tiempo  $t_3$  toma cualquier valor dentro del intervalo  $x_3 + \Delta x_3$ , si en

el momento de tiempo  $t_2$  el proceso tenía un valor dentro del intervalo  $x_2 + \Delta x_2$ , siendo  $t_1 < t_2 < t_3$ . Ambos eventos, les corresponden las siguientes dos probabilidades:

$$p(x(t_1)|x(t_2))dx_2$$

$$p(x(t_2)|x(t_3))dx_3$$

obtenidas a partir de la consideración de la condición de continuidad del proceso, y la cuales se consideran iguales dentro de los intervalos infinitamente pequeños,  $x_i + \Delta x_i$ ,  $i = 1, 2$ , de cambio del proceso.

Debido a que ambos sucesos analizados, para los procesos de Markov, se consideran independientes, entonces la probabilidad de cambio del proceso del estado  $x_1$  al estado  $x_3$ , con las condiciones de transitividad antes hechas, corresponde a la multiplicación de ambas probabilidades,

$$p(x(t_1)|x(t_2))dx_2 p(x(t_2)|x(t_3))dx_3 \quad (4.18)$$

De la última ecuación (4.23) se obtiene la probabilidad de cambio del estado  $x_1$  a  $x_3$ , teniendo en cuenta (2.39). Además, la misma probabilidad de cambio del estado  $x_1$  a  $x_3$  sin importar cual sea el valor del proceso en  $t_2$  (pero asumiendo aún la continuidad del proceso) corresponde a

$$p(x(t_1)|x(t_3))dx_3 \quad (4.19)$$

Igualando ambas expresiones (4.18) y (4.19) se obtiene la *ecuación generalizada de Markov*,

$$p(x(t_1)|x(t_3)) = \int_{-\infty}^{\infty} p(x(t_1)|x(t_2))p(x(t_2)|x(t_3))dx_2 \quad (4.20)$$

**Ejemplo 4.7.** Sea una partícula que en el momento  $k\Delta t$  salta al valor  $\pm\Delta x$ , con probabilidad  $p$  si aumenta ( $+\Delta x$ ) o con probabilidad  $q = 1 - p$  si disminuye ( $-\Delta x$ ). Basados en las relaciones obtenidas para los procesos discretos homogéneos de Markov, pero asumiendo su no estacionariedad debida al cambio en el tiempo de la distribución en el tiempo, hallar el respectivo sistema de ecuaciones diferenciales que describen el comportamiento estocástico de la partícula.

El cambio de los modelos de procesos discretos de Markov a las respectivas de los procesos continuos se realiza hallando el cambio límite de los valores  $\Delta \rightarrow 0$  y  $\Delta t \rightarrow 0$  para los modelos de ecuaciones diferenciales que los describen.

En particular, para las condiciones dadas de cambio de la partícula, la ecuación diferencial (4.15) tiene la forma,

$$P(t, x) = pP(t - \Delta t, x - \Delta t) + qP(t - \Delta t, x + \Delta t)$$

Se halla la descomposición por serie de potencias en términos de  $\Delta t$  y  $\Delta x$  de la parte derecha

de la ecuación:

$$\begin{aligned} P(t, x) = & pP(t, x) - p \frac{\partial P}{\partial t} \Delta t - p \frac{\partial P}{\partial x} \Delta x + qP(t, x) - q \frac{\partial P}{\partial t} \Delta t + \\ & + q \frac{\partial P}{\partial x} \Delta x + p \frac{\partial^2 P}{\partial t \partial x} \Delta t \Delta x + \frac{p}{2} \frac{\partial^2 P}{\partial t^2} \Delta t^2 + \frac{p}{2} \frac{\partial^2 P}{\partial x^2} \Delta x^2 - \\ & - q \frac{\partial^2 P}{\partial t \partial x} \Delta t \Delta x + \frac{q}{2} \frac{\partial^2 P}{\partial t^2} \Delta t^2 + \frac{q}{2} \frac{\partial^2 P}{\partial x^2} \Delta x^2 + \dots \end{aligned}$$

Debido a que  $p + q = 1$ , entonces se obtienen los primeros términos para la serie anterior:

$$0 = -\frac{\partial P}{\partial t} \Delta t - (p - q) \frac{\partial P}{\partial x} \Delta x + (p - q) \frac{\partial^2 P}{\partial t \partial x} \Delta t \Delta x + \frac{1}{2} \frac{\partial^2 P}{\partial t^2} \Delta t^2 + \frac{1}{2} \frac{\partial^2 P}{\partial x^2} \Delta x^2 + \dots$$

Luego, se dividen todos los términos de la serie de representación por  $\Delta t$ ,

$$\frac{\partial P}{\partial t} = -(p - q) \frac{\partial P}{\partial x} \frac{\Delta x}{\Delta t} + (p - q) \frac{\partial^2 P}{\partial t \partial x} \Delta x + \frac{1}{2} \frac{\partial^2 P}{\partial t^2} \Delta t + \frac{1}{2} \frac{\partial^2 P}{\partial x^2} \frac{\Delta x^2}{\Delta t} + \dots$$

Seguidamente, se halla el límite de la última expresión para  $\Delta t \rightarrow 0$ . Si se asume que la probabilidad de aparición de los valores grandes  $\Delta x$  decrece al disminuir el intervalo de tiempo  $\Delta t$ , tan rápido que todos los momentos de la diferencia  $\Delta x$ , comenzando desde el tercer discreto de tiempo tienden a cero más rápido que  $\Delta t$ .

De lo anterior, se obtiene los siguientes valores límite para cada uno de los términos de la serie:

1.  $\lim_{\Delta t \rightarrow 0} (p - q) \frac{\Delta x}{\Delta t} = a$
2.  $\lim_{\substack{\Delta t \rightarrow 0 \\ \Delta x \rightarrow 0 \\ (p-q) \rightarrow 0}} (p - q) \Delta x = 0$
3.  $\lim_{\Delta t \rightarrow 0} \frac{1}{2} \frac{\partial^2 P}{\partial t^2} \Delta t = 0,$
4.  $\lim_{\Delta t \rightarrow 0} \frac{\Delta x^2}{\Delta t} = b$

Por lo tanto, la operación de límite, reemplazando la probabilidad por la respectiva función densidad de probabilidad, resulta en la ecuación,

$$\frac{\partial p}{\partial t} = -a \frac{\partial p}{\partial x} + \frac{b}{2} \frac{\partial^2 p}{\partial x^2} \quad (1)$$

La solución de (1), teniendo en cuenta la condición de unidad de área,  $\int p(x, t) dx = 1$ , además, asumiendo las condiciones iniciales  $x_0$ , dadas en  $t_0$ , se demuestra que corresponde a la FDP Gaussiana,

$$p(x, t) = \frac{1}{\sqrt{2\pi b(t - t_0)}} \exp \left( -\frac{((x - x_0) - a(t - t_0))^2}{2b(t - t_0)} \right)$$

De la solución se observa que el coeficiente  $a$  corresponde al valor medio del tiempo o la velocidad de cambio del valor medio, mientras  $b$  es la varianza de los valores del tiempo.

En general, la densidad de probabilidad  $p(x(t_1)|x(t_3))$ , que cambia en función de los parámetros iniciales  $(x_1, t_1)$ , cumple la ecuación diferencial,

$$\frac{\partial p}{\partial t} + a(x, t) \frac{\partial p}{\partial x} + \frac{1}{2} b(t, x) \frac{\partial^2 p}{\partial t^2} = 0 \quad (4.21)$$

donde  $a(x, t)$  y  $b(t, x)$  pueden ser funciones variantes en el tiempo, que determinan las particularidades del proceso de Markov, y que corresponden a los siguientes momentos:

$$\begin{cases} a(x, t) &= \lim_{\Delta t \rightarrow 0} \frac{1}{\Delta t} \int_{-\infty}^{\infty} (x_2 - x_1) p(x(t_1)|x(t_1 + \Delta t)) dx_2 = \lim_{\Delta t \rightarrow 0} \frac{1}{\Delta t} E\{\Delta x|x\} \\ b(x, t) &= \lim_{\Delta t \rightarrow 0} \frac{1}{\Delta t} \int_{-\infty}^{\infty} (x_2 - x_1)^2 p(x(t_1)|x(t_1 + \Delta t)) dx_2 = \lim_{\Delta t \rightarrow 0} \frac{1}{\Delta t} E\{\Delta x^2|x\} \end{cases} \quad (4.22)$$

**Ejemplo 4.8.** Sea un proceso aleatorio al cual le corresponde el modelo en forma de la ecuación diferencial, que describe el movimiento Browniano,

$$\frac{d\xi}{dt} + \alpha \xi(t) = m\eta(t) \quad (1)$$

donde  $\alpha$  y  $m$  son constantes,  $\eta(t)$  es ruido blanco Gaussiano con valor medio igual a cero y función de correlación  $R_\xi(\tau) = \delta(\tau)$ . Hallar los momentos de la FDP del respectivo modelo de Markov.

Al asumir la total independencia de los valores del proceso  $\xi(t)$ , del modelo propuesto resulta que  $\eta(t)$  es la solución de una ecuación de primer orden, la cual se determina por las condiciones iniciales del proceso de manera unívoca. Además, los valores de la señal  $\eta(t)$  son independientes, cuyos valores pasados no influyen en los futuros. Por lo tanto, el proceso  $\xi(t)$  se puede considerar del tipo Markov y para la densidad de probabilidad condicional se cumple la ecuación (4.21). Los respectivos coeficientes de la ecuación se hallan integrando la ecuación diferencial del modelo (1) dentro de los límites  $(t, t + \Delta t)$ , con lo cual se obtiene,

$$\Delta x = m \int_t^{t+\Delta t} \eta(\lambda) d\lambda - \alpha \int_t^{t+\Delta t} \xi(\lambda) d\lambda$$

entonces,

$$E\{\Delta x\} = E\left\{\int_t^{t+\Delta t} \eta(\lambda) d\lambda\right\} - \alpha E\left\{\int_t^{t+\Delta t} \xi(\lambda) d\lambda\right\} = -\alpha \int_t^{t+\Delta t} E\{\xi(\lambda)\} d\lambda = -\alpha x \Delta t$$

De manera similar, se obtiene el segundo momento,

$$\begin{aligned} E\{\Delta t^2|x\} &= E\{m^2 \int_t^{t+\Delta t} \int_t^{t+\Delta t} \eta(\lambda_1) \eta(\lambda_2) d\lambda_1 d\lambda_2\} = m^2 \int_t^{t+\Delta t} \int_t^{t+\Delta t} E\{\eta(\lambda_1) \eta(\lambda_2)\} d\lambda_1 d\lambda_2 \\ &= m^2 \int_t^{t+\Delta t} \int_t^{t+\Delta t} R_\xi(\lambda_2 - \lambda_1) d\lambda_1 d\lambda_2 = m^2 \int_t^{t+\Delta t} \int_t^{t+\Delta t} \delta(\lambda_2 - \lambda_1) d\lambda_1 d\lambda_2 = m^2 \Delta t \end{aligned}$$

De las anteriores relaciones, teniendo en cuenta las respectivas definiciones (4.22), resultan los siguientes coeficientes  $a(x, t) = -\alpha x$  y  $b(x, t) = m^2$ .

**Ejemplo 4.9.** Hallar la solución de la ecuación diferencial, obtenida en el ejemplo 4.8, para la FDP condicional del proceso  $\xi(t)$  en el momento  $\tau$  con condiciones iniciales  $\xi(0) = x$ .

La solución de la ecuación 4.21, en general, depende de la forma de cambio de las funciones coeficiente (4.22). Una aproximación común supone la siguiente relación lineal de ambos coeficientes (pero asumiendo su estacionariedad):  $a(t, x) = \alpha_0 + \alpha_1 x$  y  $b(t, x) = \beta_0$ , donde  $\alpha_i, \beta_i, i = 1, 2$ , constantes. En este caso, es preferible el análisis realizarlo, no con la FDP, sino mediante la respectiva función característica, descrita en (2.17), con lo cual se obtiene la siguiente ecuación en derivadas parciales de primer orden:

$$\frac{\partial \Theta}{\partial t} + \left( \frac{1}{2} \beta_0 \omega^2 - j \alpha_0 \omega \right) - \alpha_1 \omega \frac{\partial}{\partial \omega} (\ln \Theta) = 0$$

que tiene solución,

$$\ln \Theta = -\frac{\sigma_\xi^2(t)}{2} \omega^2 + j \omega \overline{\xi(t)}$$

con coeficientes,

$$\begin{cases} \overline{\xi(t)} = \frac{\alpha_0}{\alpha_1} (e^{\alpha_1 t} - 1) + x e^{\alpha_1 t} \\ \sigma_\xi^2(t) = \frac{\beta_0}{2\alpha_1} (e^{2\alpha_1 t} - 1) \end{cases} \quad (2)$$

para las condiciones iniciales:  $\sigma_\xi(t) = 0, \overline{\xi(0)} = x$ .

De acuerdo con los resultados obtenidos en el ejemplo 4.8, entonces  $\alpha_0 = 0, \alpha_1 = -\alpha$  y  $\beta_0 = m^2$ , cuyos valores se reemplazan en la ecuación (2), luego, se obtiene que el proceso  $\xi(t)$  es de estructura Gaussiana con momentos,

$$\overline{\xi(t)} = x \exp(-\alpha t), \quad (3a)$$

$$\sigma_\xi^2(t) = \frac{m^2}{2\alpha} (1 - \exp(2\alpha t)) \quad (3b)$$

La varianza (3b) corresponde a la expresión de la varianza que se obtiene en el ejemplo 5.3 para el RBG a la salida del circuito  $RC$ , descrito por una ecuación diferencial con modelo similar a (1) en el ejemplo 4.8. En otras palabras, tanto el análisis de correlación, como el análisis de modelos de Markov convergen, sin embargo, en el último caso, se puede hacer una interpretación más profunda de los resultados, en particular, se puede afirmar que un proceso de Markov, con función de correlación  $R_\eta(t, \tau) = \frac{m^2}{2\alpha} e^{\alpha|\tau|}$ , también es de estructura Gaussiana. Viceversa, si un proceso Gaussiano estacionario tiene función de correlación  $R_\eta(t, \tau) = \frac{m^2}{2\alpha} e^{\alpha|\tau|}$ , entonces también es del tipo Markoviano.

**Descripción generalizada.** La descripción general de la dinámica estadística de un proceso continuo aleatorio  $x(t)$  con dimensión  $n$  en el tiempo se da por la ecuación diferencial:

$$\frac{d}{dt} x(t) = f(x, t) + \xi(x, t), \quad x(t_0) = x_0 \quad (4.23)$$

donde  $f(x, t)$  es el vector función no aleatorio con dimensión  $n$  y  $\xi(x, t)$  es el proceso aleatorio con propiedades probabilísticas conocidas, que en principio pueden depender del

vector  $\mathbf{x}$ . La expresión (4.23) se describe por medio de diferenciales en la forma:

$$d\mathbf{x}(t) = \mathbf{f}(\mathbf{x}, t)dt + d\boldsymbol{\eta}(\mathbf{x}, t), \quad \mathbf{x}(t_0) = \mathbf{x}_0 \quad (4.24)$$

donde  $d\boldsymbol{\eta}$  es el diferencial de algún proceso aleatorio  $d\boldsymbol{\eta}(\mathbf{x}, t)$  relacionado con el proceso aleatorio  $\boldsymbol{\xi}(\mathbf{x}, t)$ . Se puede demostrar [30], que siendo solución de la anterior ecuación diferencial, el proceso  $\mathbf{x}(t)$  es del tipo Markoviano, si al mismo tiempo  $\boldsymbol{\eta}(\mathbf{x}, t)$  es un proceso con incrementos (de cualquier tipo, positivos o negativos),  $\Delta\boldsymbol{\eta}(t_k) = \boldsymbol{\eta}(t_{k+1}) - \boldsymbol{\eta}(t_k)$ , que sean estadísticamente independientes. El proceso con incrementos independientes es del tipo homogéneo si se cumple que  $\Delta\boldsymbol{\eta}(t_k, t) = \Delta\boldsymbol{\eta}(t_k)$ , para los cuales la FDP es infinitamente divisible, lo que significa que si un intervalo de tiempo  $[t_0, t]$  se divide en  $N$  segmentos de igual duración  $\Delta t$ , entonces, para la función característica de los incrementos del proceso  $\boldsymbol{\eta}$  tiene lugar la igualdad

$$\Theta(j\boldsymbol{\vartheta}, t - t_0) = (\Theta(j\boldsymbol{\vartheta}, \Delta t))^N$$

En general, cualquier variable aleatoria, con FDP infinitamente divisible, se puede representar en forma de una suma de dos componentes independientes: una con estructura Gaussiana y la otra con estructura de Poisson,

$$\ln \Theta(j\boldsymbol{\vartheta}, \Delta t) = \Delta t \left( j\boldsymbol{\vartheta}^\top \mathbf{m}_1 - \frac{1}{2} \boldsymbol{\vartheta}^\top \mathbf{Q} \boldsymbol{\vartheta} + \int \left( e^{j\boldsymbol{\vartheta}^\top \mathbf{y}} - 1 \right) \pi(\mathbf{y}) d\mathbf{y} \right)$$

La primera componente,  $\Delta\dot{\mathbf{v}}$ , se puede analizar como la derivada de un proceso continuo Gaussiano, descrito tanto por el vector  $\mathbf{m}_1 \Delta t$  de valores medios con dimensión  $\mathbf{n}$  como por la matriz  $\mathbf{Q}_{\mathbf{n} \times \mathbf{n}} \Delta t$  de varianzas, que corresponde a un proceso  $\mathbf{v}(t)$  del tipo Wiener. Asumiendo  $\mathbf{m}_1 = 0$ , sin pérdida de generalidad, entonces, la componente Gaussiana se puede escribir en la forma:

$$\Delta\dot{\mathbf{v}}(t) = \mathbf{G} \Delta\mathbf{v}(t)$$

siendo  $\mathbf{G}_{\mathbf{n} \times \mathbf{k}}$ , tal que  $\mathbf{G} \mathbf{G}^\top = \mathbf{Q}$ , mientras  $\mathbf{v}(t)$  es el proceso Wiener con dimensión  $\mathbf{k}$ , de valor medio cero y matriz de varianza  $\mathbf{I} \Delta t$ .

La componente de Poisson,  $\Delta\boldsymbol{\gamma}(t)$ , que corresponde a los incrementos independientes del proceso, se puede representar como:

$$\Delta\boldsymbol{\gamma}(t) = \int \mathbf{C}(\mathbf{y}, t) \nu(d\mathbf{y}, \Delta t)$$

donde  $\mathbf{C}(\mathbf{y}, t)$  es una función determinada con dimensión  $\mathbf{n}$ , que corresponde a los valores de amplitud de los cambios o saltos que pueda tomar el proceso  $\Delta\boldsymbol{\gamma}(t)$ ;  $\nu(\mathbf{A}, \Delta t)$  es una variable aleatoria, que toma los valores  $0, 1, \dots, \mathbf{k}$  con FDP Poisson:

$$P(\nu = \mathbf{k}) = \frac{(\lambda(\mathbf{A}) \Delta t)^\mathbf{k}}{\mathbf{k}!} \exp(-\lambda(\mathbf{A}) \Delta t), \quad \lambda(\mathbf{A}) = \int_{\mathbf{A}} \pi(\mathbf{y}) d\mathbf{y}$$

Se puede demostrar que la componente Gaussiana del proceso con incrementos independientes es del tipo *estocástico continuo*, esto es, se cumple la relación,

$$\lim_{\Delta t \rightarrow 0} P(\|\eta(t + \Delta t) - \eta(t)\| > \varepsilon) = 0, \quad \forall \varepsilon > 0$$

Mientras, que para la componente de Poisson esta relación no se cumple (*Proceso estocástico con incrementos*).

A partir de lo anterior, se obtiene la descripción generalizada de la dinámica de los procesos continuos de Markov, por medio de las *ecuaciones estocásticas diferenciales*:

$$\frac{d}{dt} \mathbf{x}(t) = \mathbf{f}(\mathbf{x}, t) + \mathbf{G}(\mathbf{x}, t) \mathbf{N}(t) + \mathbf{r}(t) \quad (4.25a)$$

$$\begin{aligned} \dot{\mathbf{x}}(t) &= \mathbf{f}(\mathbf{x}, t) dt + \mathbf{G}(\mathbf{x}, t) d\mathbf{v}(t) + \dot{\mathbf{r}}(t) \\ &= \mathbf{f}(\mathbf{x}, t) dt + \mathbf{G}(\mathbf{x}, t) d\mathbf{v}(t) + \int \mathbf{C}(\mathbf{x}, \mathbf{y}, t) \nu(d\mathbf{y}, dt) \end{aligned} \quad (4.25b)$$

donde  $\nu(\mathbf{A}, \Delta t)$  se caracteriza por la función  $\pi(\mathbf{y} | t, \mathbf{x})$ , mientras,  $\mathbf{r}(t) = \dot{\mathbf{r}}(t)$  es el proceso, que corresponde la derivada en el tiempo de la componente de Poisson, y la cual se puede interpretar como una serie de pulsos  $\delta$ :

$$\mathbf{r}(t) = \sum_k \mathbf{C}(\mathbf{y}_k) \delta(t - t_k) \quad (4.26)$$

siendo  $\{t_k\}$ , el conjunto de valores aleatorios de los instantes de tiempo en los que ocurren los pulsos. La ecuaciones (4.25a) y (4.25b) se entienden en el *sentido de Stratanovich* [31], en las que aparecen el proceso aleatorio  $\mathbf{N}(t)$  con varianza infinita y un proceso  $\mathbf{r}(t)$ , el cual toma valores infinitos en los momentos de definición  $t = t_k$ , a diferencia de lo que ocurre en una ecuación diferencial ordinaria de la forma  $\dot{\mathbf{x}}(t) = \mathbf{f}(\mathbf{x}, t)$ , con condición inicial  $\mathbf{x}(t_0) = \mathbf{x}_0$ , y para la cual se supone la solución,

$$\mathbf{x}(t) = \mathbf{x}_0 + \int_{t_0}^t \mathbf{f}(\mathbf{x}, \tau) d\tau$$

entonces, la integral se entiende en el *sentido de Cauchy-Riemann* y se da por el límite de la suma, el cual existe si la función  $\mathbf{f}(\mathbf{x}, \tau)$  cumple las condiciones de Lipschitz [1].

$$\int_{t_0}^t \mathbf{f}(\mathbf{x}, \tau) d\tau = \lim_{\Delta t \rightarrow 0} \sum_{k=0}^{N-1} \mathbf{f}(\mathbf{x}(t_i), t_i) \Delta t, \quad \Delta t = t_{k+1} - t_k$$

De manera similar, se asume la solución del ecuación diferencial estocástica (4.25b), concretamente en la forma:

$$\mathbf{x}(t) = \mathbf{x}_0 + \int_{t_0}^t \mathbf{f}(\mathbf{x}, \tau) d\tau + \int_{t_0}^t \mathbf{G}(\mathbf{x}, \tau) d\mathbf{v}(\tau) + \int_{t_0}^t \int \mathbf{C}(\mathbf{x}, \mathbf{y}, \tau) \nu(d\mathbf{y}, d\tau) \quad (4.27)$$

para la cual existen, por lo menos, dos formas diferentes de definición de la integral de los procesos aleatorios. La primera integral se puede definir, en analogía con la forma Cauchy-Riemann, si se asume su convergencia en el sentido del valor cuadrático medio, esto es, l.i.m.  $\xi_n = \xi$ , cuando se cumple  $\lim_{n \rightarrow \infty} E \left\{ (\xi_n - \xi)(\xi_n - \xi)^T \right\} = 0$ .

La integral de la componente de Poisson, en correspondencia con (4.26), se escribe como:

$$\int_{t_0}^t \int \mathbf{C}(\mathbf{x}, \mathbf{y}, \tau) \nu(d\mathbf{y}, d\tau) = \int_{t_0}^t \sum_k \mathbf{C}(\mathbf{x}, \mathbf{y}, \tau) \delta(\tau - t_k) d\tau$$

En cuanto a la integral de la componente Gaussiana de (4.27), ésta se puede definir de diferente manera. La primera forma corresponde a la *integral estocástica en el sentido de Ito*, entendida como la suma del límite de las sumas convergentes en el valor cuadrático medio del tipo:

$$\hat{\mathbf{S}}(t) = \int_{t_0}^t \mathbf{G}(\mathbf{x}, \tau) d\hat{\nu}(\tau) = \lim_{N \rightarrow \infty} \sum_{k=0}^{N-1} \mathbf{G}(\mathbf{x}(t_k, t_k)) (\nu(t_{k+1}) - \nu(t_k))$$

La ecuación diferencial, para la cual la correspondiente integral se entiende en el sentido de Ito, se denomina *ecuación estocástica de Ito*, descrita como:

$$d\mathbf{x}(t) = \mathbf{A}(\mathbf{x}, t)dt + \mathbf{G}(\mathbf{x}, t)d\hat{\nu}(t) + \gamma(t) \quad (4.28a)$$

$$\frac{d}{dt}\mathbf{x}(t) = \mathbf{A}(\mathbf{x}, t) + \mathbf{G}(\mathbf{x}, t)\hat{\mathbf{N}}(t) + \mathbf{r}(t) \quad (4.28b)$$

Si un proceso de Markov,  $\mathbf{x}(t)$ , con dimensión  $n$  cumple la condición (4.28a), entonces el proceso  $\mathbf{z}(t) = \Phi(\mathbf{x}, t)$ , siendo  $\Phi(\mathbf{x}, t)$  cierto vector funcional determinado con dimensión  $L$ , cumple la ecuación conocida como la *expresión de Ito*:

$$\begin{aligned} dz(t) = & \left\{ \frac{\partial}{\partial t} \Phi(\mathbf{x}, t) + \left( \frac{\partial}{\partial \mathbf{x}} \Phi(\mathbf{x}, t) \right) \mathbf{A}(\mathbf{x}, t) + \right. \\ & \left. \frac{1}{2} \left( \sum_m^N \sum_n^N \sum_l^K \frac{\partial^2 \Phi_i(\mathbf{x}, t)}{\partial x_m \partial x_n} G_{ml}(\mathbf{x}, t) G_{nl}(\mathbf{x}, t) \right)_{i=1, \dots, L} \right\} dt + \\ & + \left\{ \frac{\partial}{\partial \mathbf{x}} \Phi(\mathbf{x}, t) \right\} \mathbf{G}(\mathbf{x}, t) d\nu(t) + \\ & + \int (\Phi(\mathbf{x} + \mathbf{C}(\mathbf{x}, \mathbf{y}, t), t) - \Phi(\mathbf{x}, t)) \nu(d\mathbf{y}, dt) \end{aligned}$$

donde  $\Phi_i(\mathbf{x}, t)$ ,  $i = 1, \dots, L$  son las componentes del vector  $\Phi(\mathbf{x}, t)$ ,  $G_{ml}(\mathbf{x}, t)$  y  $G_{nl}(\mathbf{x}, t)$  son las componentes de la matriz  $\mathbf{G}(\mathbf{x}, t)$ .



La expresión de Ito se puede simplificar en la forma

$$dz(t) = \frac{\partial \Phi}{\partial t} dt + \frac{\partial \Phi}{\partial x} dx(t) + \frac{1}{2} \left( \sum_{m,n=1}^N \sum_{l=1}^K \frac{\partial^2 \Phi_i}{\partial x_m \partial x_n} G_{ml} G_{nl} \right)_{i=1,\dots,L} dt + \int \left( \Phi(x + C(x, y, t), t) - \Phi(x, t) - \frac{\partial \Phi}{\partial x} C(x, y, t) \right) \nu(dy, dt)$$

de la que se observa que si un proceso de Markov  $x(t)$  es continuo ( $C(x, y, t) = 0$ ), entonces el diferencial del proceso  $z(t)$ , con precisión hasta el término no aleatorio, es proporcional al diferencial del proceso  $x(t)$ . A propósito, el tercer término de la última ecuación, hace imposible el cambio de variables en forma convencional, integrando por partes, o mediante otras técnicas de integración. Además, la integral estocástica de Ito, determinada para el sentido directo del tiempo, no coincide con la de tiempo inverso, cuando en la definición de  $\hat{S}$  la función  $G(x, t)$  se toma en el extremo opuesto del intervalo elemental.

En [31] se define la integral estocástica en el sentido de Stratanovich, que es relativamente simétrica con respecto al pasado y futuro, y que corresponde al límite de las sumas convergentes en el valor cuadrático medio de la forma:

$$S(t) = \int_{t_0}^t G(x, \tau) d\nu(\tau) = \lim_{N \rightarrow \infty} \sum_{i=0}^{N-1} G\left(\frac{x(t_{i+1}) + x(t_i)}{2}, t_i\right) (\nu(t_{i+1}) - \nu(t_i))$$

En este caso, para la formación de las sumas integrales, los valores de  $G(X, \tau)$  se toman en los puntos medios de los subintervalos elementales. Una de las propiedades de la integral estocástica de Stratanovich está en la posibilidad de su empleo en las formas convencionales del análisis matemático. Las correspondientes ecuaciones (4.25a) y (4.25b), también se denominan *ecuaciones estocásticas en forma simétrica*. La integral simétrica se diferencia de la respectiva integral de Ito en el valor [31]:

$$S(t) - \hat{S}(t) = \frac{1}{2} \int_{t_0}^t \left( \sum_{m=1}^N \sum_{l=1}^K \frac{\partial G_{il}(x, \tau)}{\partial x_m} G_{ml}(x, \tau) \right) d\tau, \quad i = 1, \dots, N$$

esto es, exceptuando el caso cuando  $G(x, \tau)$  no depende de  $x$ , ambos integrales se diferencian, por esto es importante en la descripción de las ecuaciones estocásticas diferenciales, determinar en que forma se deben entender las ecuaciones. De otra parte, de la última ecuación resulta que las expresiones (4.25a) y (4.25b) en el sentido de Stratanovich y las respectivas (4.28a) y (4.28b) en el sentido de Ito, describen un mismo proceso de Markov, cuando se cumple la igualdad,

$$A_i(x, t) = f_i(x, t) + \frac{1}{2} \sum_{m=1}^N \sum_{l=1}^K \frac{\partial G_{il}(x, t)}{\partial x_m} G_{ml}(x, t) \quad (4.29)$$

donde  $A_i(x, t)$  y  $f_i(x, t)$  son las respectivas componentes de los vectores funciones.

La descripción de los procesos continuos de Markov incluye la definición de la FDP de transición  $\pi(\mathbf{x}, t \mid \mathbf{x}_0, t_0)$  y la FDP marginal  $p(\mathbf{x}, t)$ , relacionadas por la expresión:

$$p(\mathbf{x}, t) = \int_{\mathbf{x} \in X} \pi(\mathbf{x}, t \mid \mathbf{x}_0, t_0) p(\mathbf{x}_0, t_0) d\mathbf{x}_0 \quad (4.30)$$

que cumplen las condiciones  $\int_{\mathbf{x} \in X} p(\mathbf{x}, t) d\mathbf{x} = 1$ ,  $\int_{\mathbf{x} \in X} \pi(\mathbf{x}, t \mid \mathbf{x}_0, t_0) d\mathbf{x} = 1$ .

Además, para los procesos continuos de Markov es cierta la *ecuación de Kolmogorov-Chapman*:

$$\pi(\mathbf{x}, t \mid \mathbf{x}_0, t_0) = \int_{\mathbf{x} \in X} \pi(\mathbf{x}, t \mid \mathbf{x}_1, t_1) \pi(\mathbf{x}_1, t_1 \mid \mathbf{x}_0, t_0) d\mathbf{x}_1$$

donde  $X$  es el dominio de los valores del proceso  $\mathbf{x}(t)$ ,  $t > t_1, t_0$ .

Se puede demostrar [32], que si todos los elementos del vector  $\mathbf{A}(\mathbf{x}, t)$  son diferenciables, así mismo los elementos de  $\mathbf{G}(\mathbf{x}, t)$  son dos veces diferenciables por  $\mathbf{x}$ , entonces la FDP de transición  $\pi(\mathbf{x}, t \mid \mathbf{x}_0, t_0)$ , dada en función de  $\mathbf{x}$  y  $t$ , cumple la *ecuación directa de Kolmogorov-Feller*:

$$\frac{\partial}{\partial t} \pi(\mathbf{x}, t \mid \mathbf{x}_0, t_0) = \mathcal{K}_{t,x} \{ \pi(\mathbf{x}, t \mid \mathbf{x}_0, t_0) \} \quad (4.31)$$

con condiciones iniciales  $\pi(\mathbf{x}, t \mid \mathbf{x}_0, t_0) = \delta(\mathbf{x} - \mathbf{x}_0)$ . El operador  $\mathcal{K}_{t,x} \{ \cdot \}$ , denominado *primer operador derivante* se determina por la suma de los operadores

$$\mathcal{K}_{t,x} \{ \cdot \} = \mathcal{L}_{t,x} \{ \cdot \} + \mathcal{M}_{t,x} \{ \cdot \}$$

definidos por las respectivas expresiones:

$$\begin{aligned} \mathcal{L}_{t,x} \{ w(\mathbf{x}) \} &= - \sum_{i=1}^n \frac{\partial}{\partial x_i} (A_i(\mathbf{x}, t) w(\mathbf{x})) + \frac{1}{2} \sum_{i=1}^N \sum_{l=1}^N \frac{\partial^2}{\partial x_i \partial x_l} (B_{il}(\mathbf{x}, t) w(\mathbf{x})) \\ \mathcal{M}_{t,x} \{ w(\mathbf{x}) \} &= \int \int w(\xi) (\delta(\mathbf{x} - \xi - \mathbf{C}(\xi, \mathbf{y}, t)) - \delta(\mathbf{x} - \xi)) \pi(\mathbf{y} \mid t, \xi) d\mathbf{y} d\xi \end{aligned}$$

donde  $A_i(\mathbf{x}, t)$  se denominan la *deriva del proceso* y se determinan de la expresión (4.29), mientras  $B_{il}(\mathbf{x}, t)$  son elementos positivos que determinan la matriz denominada *matriz de difusión* de orden  $n \times n$  e igual a

$$\mathbf{B}(\mathbf{x}, t) = \mathbf{G}(\mathbf{x}, t) \mathbf{G}^T(\mathbf{x}, t)$$

La FDP de transición, como función de  $\mathbf{x}_0$  y  $t_0$ , cumple con la *ecuación inversa de Kolmogorov-Feller*,

$$-\frac{\partial}{\partial t} \pi(\mathbf{x}, t \mid \mathbf{x}_0, t_0) = \mathcal{K}_{t,x} \{ \pi(\mathbf{x}, t \mid \mathbf{x}_0, t_0) \} \quad (4.32)$$

mediante el operador inverso  $\mathcal{K}_{t,x}^{-1} \{\cdot\} = \mathcal{K}_{t,x}^* \{\cdot\}$ , que resulta ser su propio conjugado:

$$\begin{aligned}\mathcal{L}_{t,x}^* \{w(x)\} &= \sum_{i=1}^n A_i(x, t) \frac{\partial w(x)}{\partial x_i} + \frac{1}{2} \sum_{i=1}^N \sum_{l=1}^N B_{il}(x, t) \frac{\partial^2 w(x)}{\partial x_i \partial x_l} \\ \mathcal{M}_{t,x}^* \{w(x)\} &= \int (w(x + \mathbf{C}(x, \mathbf{y}, t)) - w(x)) \pi(\mathbf{y} | t, x) d\mathbf{y}\end{aligned}$$

Debido a la relación entre los operadores directo e inverso, se tiene que para cualquier dominio  $\Omega$ , en el cual los elementos de la función  $\mathbf{A}(x, t)$  son diferenciables, además los elementos de  $\mathbf{B}(x, t)$  son dos veces diferenciables por  $x$ , es cierta la relación:

$$\int_{\Omega} u(x) \mathcal{K}_{t,x} \{w(x)\} dx = \int_{\Omega} w(x) \mathcal{K}_{t,x}^* \{u(x)\} dx$$

donde  $u(x)$  y  $w(x)$  son funciones escalares, de las cuales por lo menos una con sus primeras derivadas son iguales a cero en la frontera del dominio  $\Omega$ .

La FDP marginal dada en (4.30), teniendo en cuenta (4.31), corresponde a la ecuación

$$\frac{\partial}{\partial t} p(x, t) = \mathcal{K}_{t,x} \{p(x, t)\} \quad (4.33)$$

que se debe resolver para la condición inicial  $p(x, t) = p(x_0)$ .

Las ecuaciones (4.31), (4.32) y (4.33) son del tipo integro-diferenciales con derivadas parciales, para las cuales no hay una solución compacta e inclusive para el caso estacionario la solución es muy compleja.

Si no existe la componente de Poisson en un proceso con incrementos independientes en la ecuaciones diferenciales estocásticas (4.25b) y (4.28a), entonces el proceso continuo de Markov  $x(t)$  se denomina *difusivo*, para el cual las ecuaciones (4.31), (4.32) y (4.33) se denominan ecuaciones de *Fokker-Plank-Kolmogorov*. En los procesos difusivos de Markov se introduce el concepto de *flujo de la densidad de probabilidad*  $\Pi(x, t)$ , cuyas componentes en cualquier punto de  $x$  son iguales a

$$\Pi_i(x, t) = A_i(x, t) + p(x, t) - \frac{1}{2} \sum_{l=1}^N \frac{\partial}{\partial x_l} (B_{il}(x, t) p(x, t)), \quad i = 1, \dots, N \quad (4.34)$$

La ecuación directa de Fokker-Plank-Kolmogorov (4.34) usualmente se escribe como:

$$\frac{\partial}{\partial t} p(x, t) = \mathcal{L}_{t,x} \{p(x, t)\} = - \sum_{i=1}^N \frac{\partial}{\partial x_i} \Pi_i(x, t) = -\text{div} \Pi(x, t)$$

la cual se interpreta como la ley de conservación de la función densidad de probabilidad.

En forma general, se consideran diferentes tipos de comportamiento de un proceso difusivo de Markov, con respecto a las partículas que se difunden en la frontera del dominio, entre otros, los siguientes: absorción, reflexión, salida abrupta de la frontera, parada, etc.

El cálculo numérico de las ecuaciones de Fokker-Plank-Kolmogorov se realiza empleando

diferentes métodos de aproximación, entre ellos, los de Galerkin, iterativos, esperanzas matemáticas condicionales, y en particular, el método de Monte-Carlo, que se basa en la solución numérica de las ecuaciones diferenciales estocásticas (4.25a) y (4.25b) [30].

#### 4.1.4. Sucesiones de Markov

Corresponde a sucesiones vectores de variables aleatorias,  $\mathbf{x}_i = \mathbf{x}(t_i)$ , con dimensión  $n$ , las cuales, en determinados momentos del tiempo,  $t_0 < t_1 < \dots < t_i < \dots$ , forman una escala continua de posibles valores, que cumplen las condiciones (4.1) y (4.2). La discretización de procesos continuos aleatorios con carga informativa es un ejemplo en el cual, empleando las sucesiones de Markov, se puede analizar la dinámica estadística de los diferentes sistemas de procesamiento digital de procesos aleatorios.

De la expresión (4.2), se observa que la FDP conjunta de una sucesión de Markov puede ser expresada a través de la densidad de probabilidad del estado inicial  $p(\mathbf{x}_0)$  y la densidades de probabilidad de cambio elemental  $\pi_i(\mathbf{x}_i | \mathbf{x}_{i-1})$ ,

$$p(\mathbf{x}_0, \dots, \mathbf{x}_k) = p(\mathbf{x}_0) \prod_{i=1}^k \pi_i(\mathbf{x}_i | \mathbf{x}_{i-1})$$

En analogía con las cadenas compuestas de Markov, se pueden definir las sucesiones compuesta de Markov de orden  $m > 1$ , de tal manera que se cumple

$$\pi_k(\mathbf{x}_k | \mathbf{x}_0, \dots, \mathbf{x}_{k-1}) = \pi_k(\mathbf{x}_k | \mathbf{x}_{k-m}, \dots, \mathbf{x}_{k-1})$$

donde  $k \leq m$ . También en este caso cualquier sucesión compuesta se puede representar mediante una combinación de cadenas simples.

Cualquier sucesión, escogida dentro de otra sucesión de Markov, vuelve a ser del tipo Markov, esto es, si para un momento dado  $t_k$  se analiza la malla de valores discretos tiempos  $t_{k_1} < t_{k_2} < \dots < t_{k_m}$ , entonces  $\pi(\mathbf{x}_{k_m} | \mathbf{x}_{k_1}, \dots, \mathbf{x}_{k_{m-1}}) = \pi_{k_m}(\mathbf{x}_{k_m} | \mathbf{x}_{k_{m-1}})$ .

Las densidades condicionales de probabilidad de transición para una sucesión de Markov cumplen con la ecuación de Kolmogorov-Chapman

$$\pi(\mathbf{x}_l | \mathbf{x}_i) = \int_{\mathbf{x} \in X} \pi(\mathbf{x}_l | \mathbf{x}_k) \pi(\mathbf{x}_k | \mathbf{x}_i) d\mathbf{x}_k \quad (4.35)$$

donde  $i < k < l$  y  $X$  es el dominio de los valores del proceso  $\mathbf{x}(t)$ . Si se nota por  $p_k(\mathbf{x} | \mathbf{x}_0)$  la densidad condicional de la probabilidad de la variable aleatoria  $\mathbf{x}_k$ , para un valor fijo de  $\mathbf{x}_0$  y considerando  $\pi_k(\mathbf{x} | \mathbf{z}) = \pi_k(\mathbf{x}_k | \mathbf{x}_{k-1})$ , de (4.35) se obtiene que

$$p_{k+1}(\mathbf{x} | \mathbf{x}_0) = \int_X \pi_k(\mathbf{x} | \mathbf{z}) p_k(\mathbf{z} | \mathbf{x}_0) d\mathbf{z} \quad (4.36)$$

Así, si se conocen las FDP elementales de transición  $\pi_k(\mathbf{x} | \mathbf{z})$ , entonces la relación recurrente (4.36) permite calcular las características estadísticas de una sucesión de Markov.

Una sucesión de Markov se considera *homogénea* si las FDP elementales de transición  $\pi_k(\mathbf{x} \mid \mathbf{z})$  no dependen de  $k$ , así mismo, se considera *estacionaria* si la cadena, además de ser homogénea, todos los estados  $\mathbf{x}_k$  tienen una sola FDP,  $p(\mathbf{x}) = \lim_{k \rightarrow \infty} p_k(\mathbf{x})$ , que de existir, cumple la ecuación integral:

$$p(\mathbf{x}) = \int_X \pi(\mathbf{x} \mid \mathbf{z}) p(\mathbf{z}) d\mathbf{z}, \quad \pi(\mathbf{x} \mid \mathbf{z}) = \lim_{k \rightarrow \infty} \pi_k(\mathbf{x} \mid \mathbf{z}) \quad (4.37)$$

Similar a (4.23) y (4.24), la descripción generalizada de la dinámica estadística de una sucesión de Markov  $\mathbf{x}_k = \mathbf{x}(t_k)$  con dimensión  $n$ , en tiempo discreto, se da como,

$$\mathbf{x}_{k+1} = \Phi_k(\mathbf{x}_k, \mathbf{w}_k) \quad (4.38)$$

donde  $\Phi_k(\mathbf{x}_k, \mathbf{w}_k)$  es la función vectorial que no es aleatoria con respecto a sus argumentos;  $\mathbf{w}_k = \mathbf{w}(t_k)$  son los valores independientes del proceso aleatorio discretizado con FDP conocidas  $q_k(\mathbf{w}) = q(\mathbf{w}, t_k)$ . Cabe anotar que la ecuación (4.38) determina la relación entre las variables aleatorias  $\mathbf{x}_{k+1}$  y  $\mathbf{x}_k$  para valores fijos de  $\mathbf{x}_k$ . Por esta razón, las FDP elementales de transición que participan en las ecuaciones (4.36) y de (4.37), se hallan de (4.38) empleando diferentes reglas de transformación [15].

En el análisis de las sucesiones de Markov, también se pueden dar diferentes comportamientos en las fronteras de un dominio dado  $\omega \subset X$ . Por ejemplo, si es de interés hallar la primera salida de una sucesión homogénea de Markov fuera de la frontera  $\Gamma \subset \Omega$ , entonces la probabilidad  $P(k, \mathbf{x}_0)$  de que la sucesión de Markov, comenzando de  $\mathbf{x}_0 \in \Omega$ , salga por primera vez de la frontera  $\Gamma$  en el estado de transición  $k$  se halla de la relación:

$$P(k, \mathbf{x}_0) = \int_{\Omega} (p_{k-1}(\mathbf{x} \mid \mathbf{x}_0) - p_k(\mathbf{x} \mid \mathbf{x}_0)) d\mathbf{x} \quad (4.39)$$

donde la densidad de probabilidad  $p_k(\mathbf{x} \mid \mathbf{x}_0)$  cumple la ecuación recurrente

$$p_k(\mathbf{x} \mid \mathbf{x}_0) = \int_{\Omega} \pi(\mathbf{x} \mid \mathbf{z}) p_{k-1}(\mathbf{z} \mid \mathbf{x}_0) d\mathbf{z} \quad (4.40)$$

De (4.39), la cantidad media de transiciones hasta la primera salida fuera de la frontera  $\Gamma \subset \Omega$  es

$$E\{k \mid \mathbf{x}_0\} = \sum_{k=1}^{\infty} k P(k, \mathbf{x}_0) = 1 + \int_{\Omega} \sum_{k=1}^{\infty} p_k(\mathbf{x} \mid \mathbf{x}_0) d\mathbf{x} = 1 + \int_{\Omega} P(\mathbf{x} \mid \mathbf{x}_0) d\mathbf{x}$$

siendo  $P(\mathbf{x} \mid \mathbf{x}_0) = \sum_{k=1}^{\infty} p_k(\mathbf{x} \mid \mathbf{x}_0)$ . la misma función  $P(\mathbf{x} \mid \mathbf{x}_0)$ , como se observa de (4.40) se puede hallar de la solución de la ecuación integral

$$P(\mathbf{x} \mid \mathbf{x}_0) = \pi(\mathbf{x} \mid \mathbf{x}_0) + \int_{\Omega} (\mathbf{x} \mid \mathbf{z}) P(\mathbf{z} \mid \mathbf{x}_0) d\mathbf{z}$$

El cálculo de la última ecuación integral se puede realizar empleando los métodos conocidos de solución para ecuaciones integrales homogéneas y no homogéneas de Fredholm.

## Problemas

**Problema 4.1.** Dadas las matrices de transición

$$\Pi = \begin{bmatrix} 0 & 2/3 & 1/3 \\ 1/2 & 0 & 1/2 \\ 1/3 & 2/3 & 0 \end{bmatrix}; \quad \Pi = \begin{bmatrix} 1/2 & 1/3 & 1/6 \\ 1/2 & 1/3 & 1/6 \\ 1/2 & 1/3 & 1/6 \end{bmatrix}, \quad \Pi = \begin{bmatrix} 0 & 1/2 & 1/2 \\ 1/2 & 0 & 1/2 \\ 1/2 & 1/2 & 0 \end{bmatrix}$$

Calcular las respectivas matrices de transición para  $n = 2, 3, 4, 5$  pasos.

**Problema 4.2.** Sea  $\xi_t$  el número de orden de un estado en la cadena de Markov en el instante de tiempo  $t$ ,  $P(\xi_0 = 1) = 1$ , y la matriz de probabilidades de transición tiene la forma

$$\Pi = \begin{bmatrix} 3/7 & 3/7 & 1/7 \\ 1/11 & 2/11 & 8/11 \\ 1/11 & 4/11 & 6/11 \end{bmatrix}$$

Demostrar que la sucesión derivada  $\eta_t = \begin{cases} 1, & \xi_t = 1 \\ 2, & \xi_t \neq 1 \end{cases}$  es un cadena de Markov y hallar su respectiva matriz de transición.

**Problema 4.3.** Un electrón puede encontrarse en una de tres órbitas con una probabilidad de cambio en unidad de tiempo de la órbita  $i$  a la  $j$  igual a,  $C_i \exp(-\lambda|i-j|)$ ,  $\lambda > 0$ . Hallar las probabilidades de cambio en dos unidades de tiempo y determinar la constante  $C_i$ .

**Problema 4.4.** Dada la matriz de transición

$$\begin{bmatrix} 0 & 1/2 & 1/2 \\ 1/4 & 1/4 & 1/4 \\ 1/2 & 1/2 & 0 \end{bmatrix}$$

establecer si es ergódica o no y calcular sus probabilidades finales.

**Problema 4.5.** Dada la matriz

$$\Pi = \begin{bmatrix} 1 & 0 \\ 1/3 & 2/3 \end{bmatrix}$$

Demostrar que la matriz  $\Pi^n$ , para  $n \rightarrow \infty$  no cumple el principio de ergodicidad, sin embargo, la matriz se puede considerar ergódica. Explicar porqué y hallar las probabilidades finales.

**Ejercicio en el CP 4.1.** Realizar un programa de realice, en el caso de las cadenas de Markov, los siguientes procedimientos:

1. Cálculo de las expresiones (4.7), dados los respectivos vectores y matrices iniciales.
2. Resolución del sistema algebraico de ecuaciones lineales (4.9).
3. Análisis de la convergencia de los elementos de transición en los valores de probabilidad final para el caso de la sucesiones estacionarias de Markov, por ejemplo en forma de un espectrograma.

**Problema 4.6.** Hallar la probabilidad de estado  $P_k(t)$  para un proceso homogéneo de Markov, en el cual se tiene  $\alpha_k = \alpha$  y  $\beta_k = k\beta$ ,  $k = 0, \dots, K$ . Resolver la ecuación diferencial asumiendo la ergodicidad del proceso aleatorio.

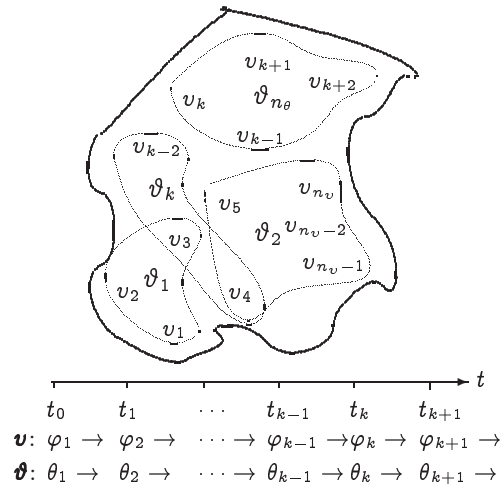
## 4.2. Análisis experimental de procesos de Markov

En el proceso real de señales, independiente de su forma de representación, análoga o discreta, no siempre está dado su modelo de aleatoriedad, y si está dado se debe resolver el problema del cálculo de sus características. La importancia de la caracterización de los modelos de las señales está en su papel básico durante la implementación de sistemas de proceso, entre ellos, los de reconocimiento, identificación, predicción, etc.

Los modelos de aleatoriedad de Markov, que se describen en el numeral §4.1, son de naturaleza probabilístico y su empleo adecuado en tareas de aplicación implica la estimación y ajuste de sus parámetros, que se realiza mediante el análisis estadístico de las señales reales. Cabe anotar, que debido a la complejidad en el análisis de los modelos de Markov, la caracterización de las señales aleatorias es preferible realizarla mediante métodos de estimación paramétrica.

### 4.2.1. Modelos ocultos de Markov

En general, las sucesiones de Markov, descritas en el numeral §4.1.4, asumen una cantidad finita de valores discretos o estados para la representación de la señal aleatoria. En particular, cada estado de manera directa se asocia a un evento físico observable. Sin embargo, en la práctica, se tienen aplicaciones con señales que no presentan de forma evidente los eventos sobre los cuales se construye el modelo. En este sentido, se debe construir un modelo probabilístico sobre los estados no observables u *ocultos*. Como resultado las cadenas construidas por este principio, corresponden a un proceso estocástico doblemente inmerso; la función probabilística de los estados ocultos y el mismo modelo de aleatoriedad de Markov impuesto sobre el proceso.



**Figura 4.4.** Representación de estados en los modelos ocultos de Markov

Los modelos ocultos de Markov, cuyos posibles estados se representan en la Figura 4.4, se pueden caracterizar mediante los siguientes parámetros:

- (a). El número de símbolos de observación o volumen  $n_v$  del alfabeto discreto de observación  $\mathbf{v} = \{v_k : k = 1, \dots, n_v\} \in \mathcal{U}$ . Los símbolos de observación corresponden a la salida física del sistema en análisis y conforman la sucesión aleatoria  $\boldsymbol{\varphi} = \{\varphi_1, \dots, \varphi_{n_\varphi}\}$  en los momentos definidos de tiempo  $1, 2, \dots, n_\varphi$ , donde  $n_\varphi$  es la longitud de la sucesión de observación.
- (b). El número de los estados ocultos del modelo,  $\boldsymbol{\vartheta} = \{\vartheta_k : k = 1, \dots, n_\vartheta\} \in \mathcal{V}$ , que siendo no observables, pueden ser relacionados con algún sentido físico del proceso. Los estados ocultos conforman la sucesión aleatoria  $\boldsymbol{\theta} = \{\theta_0, \theta_1, \dots, \theta_{n_\theta}\}$  en los momentos definidos de tiempo  $= 1, 2, \dots, n_\theta$ , donde  $n_\theta$  es la longitud de la sucesión de estados en análisis.

En general, los estados ocultos del modelo de cadenas de Markov se obtienen a partir del alfabeto de los símbolos de observación, por la transformación,  $\mathcal{V} = \mathcal{K} \{\mathcal{U}\}$ , tal que se conserven las medidas estadísticas:  $\sum_k^{n_\vartheta} P(\vartheta_k) = \sum_k^{n_v} P(v_k) = 1$ .

- (c). La matriz probabilidad de transición de estados,  $\boldsymbol{\Pi} = \{\pi_{mn} : m, n = 1, \dots, n_\vartheta\}$ , en la cual cada elemento se determina como,

$$\pi_{mn}(k) = P(\theta_{k+1} = \vartheta_n | \theta_k = \vartheta_m), \quad \pi_{mn} \geq 0, \quad \sum_{n=1}^{n_\vartheta} \pi_{mn} = 1$$

- (d). La matriz probabilidad condicional  $\mathbf{P}(\mathbf{v}|\boldsymbol{\vartheta})$  de los símbolos de observación respecto a cada estado  $\vartheta_i$  en el momento de tiempo  $n$ , con elementos,

$$P_{ki}(n) = P(v_k | \theta_n = \vartheta_i), \quad 1 \leq i \leq n_\vartheta, \quad 1 \leq k \leq n_v$$

- (e). El vector probabilidad de estado inicial  $\mathbf{p}_{\theta_1}$  con elementos  $\{P_{\theta_1}(i)\}$ , donde

$$p_{\theta_1}(i) = P(\theta_1 = \vartheta_i), \quad 1 \leq i \leq n_\vartheta$$

Los valores de aleatoriedad  $\boldsymbol{\Pi}$ ,  $\mathbf{P}(\mathbf{v}|\boldsymbol{\vartheta})$  y  $\mathbf{p}_{\theta_1}$ , notados en conjunto como

$$\boldsymbol{\mu} = \{\boldsymbol{\Pi}, \mathbf{P}(\mathbf{v}|\boldsymbol{\vartheta}), \mathbf{p}_{\theta_1}\}$$

conforman los parámetros de un modelo oculto de Markov, el cual se puede emplear para generar la estimación de la sucesión de observación,  $\boldsymbol{\varphi} \in \{\varphi_1, \varphi_2, \dots, \varphi_{n_\varphi}\}$ , con longitud  $n_\varphi = n_\vartheta$  para los momentos definidos de tiempo  $n = 1, 2, \dots, n_\varphi$ .

El desarrollo de las modelos ocultos de Markov está relacionado con las siguientes tres tareas estadísticas [33]:

1. Dada una sucesión de observación  $\boldsymbol{\varphi} = \{\varphi_1, \varphi_2, \dots, \varphi_{n_\varphi}\}$  con longitud  $n_\varphi$  y el modelo establecido,  $\boldsymbol{\mu} = \{\boldsymbol{\Pi}, \mathbf{P}(\mathbf{v}|\boldsymbol{\vartheta}), \mathbf{p}_{\theta_1}\}$ , cómo calcular de manera eficiente la probabilidad  $P(\boldsymbol{\varphi}|\boldsymbol{\mu})$  de la sucesión de observación.



2. Dada una sucesión de observación  $\boldsymbol{\varphi} = \{\varphi_1, \varphi_2, \dots, \varphi_{n_\varphi}\}$  con longitud  $n_\varphi$ , además, conocido el modelo  $\mu$ , cómo estimar de forma óptima, para una criterio de medida fijado a priori, la correspondiente sucesión de estados  $\boldsymbol{\theta} = \{\theta_1, \theta_2, \dots, \theta_{n_\varphi}\}$ .
3. El ajuste de los parámetros del modelo  $\mu$  que brinden el máximo valor de  $P(\boldsymbol{\varphi}|\mu)$ .

#### 4.2.2. Entrenamiento del modelo

**Estimación de la probabilidad  $P(\boldsymbol{\varphi}|\mu)$ .** Un primer método directo de cálculo de la probabilidad, para una sucesión de observación dada, consiste en la enumeración de cada una de los posibles sucesiones de estados que explican esa sucesión de observación.

Sea la sucesión  $i$  de estados ocultos con longitud  $n_\varphi$ ,  $\boldsymbol{\theta}(i) = \{\theta_1, \theta_2, \dots, \theta_n, \dots, \theta_{n_\varphi}\}$ , para la cual la probabilidad de la sucesión de observación, asumiendo la independencia de las observaciones, así como la independencia en el tiempo de las probabilidades  $\pi_{mn}$  y  $P_{ki}$ , está dada como

$$P(\boldsymbol{\varphi}|\boldsymbol{\theta}(i), \mu) = \prod_{n=1}^{n_\varphi} P(\varphi_n|\theta_n, \mu)$$

La probabilidad de que la sucesión dada de estados  $\boldsymbol{\theta}(i)$  ocurra se determina como,

$$P(\boldsymbol{\theta}(i)|\mu) = p_{\theta_1}(\theta_1) \pi_{\theta_1\theta_2} \cdots \pi_{\theta_{n_\varphi-1}\theta_{n_\varphi}}$$

de tal manera, que la probabilidad conjunta de que se tengan en forma simultanea ambas sucesiones corresponde al producto de los anteriores dos términos,

$$P(\boldsymbol{\varphi}, \boldsymbol{\theta}(i)|\mu) = P(\boldsymbol{\varphi}|\boldsymbol{\theta}(i), \mu)P(\boldsymbol{\theta}(i)|\mu) \quad (4.41)$$

La interpretación de la expresión (4.41) corresponde a que en el momento inicial de tiempo,  $n = 1$ , se tiene el estado  $\theta_1$  con una probabilidad de  $p_{\theta_1}(\theta_1)$  y se genera el símbolo de observación  $\varphi_1$  con una probabilidad de  $P_{\theta_1\varphi_1}$ . En el siguiente momento del tiempo, cuando  $n = 2$ , se realiza la transición al estado  $\theta_2$  desde  $\theta_1$  con una probabilidad de  $\pi_{\theta_1\theta_2}$  y se genera el símbolo  $\varphi_2$  con una probabilidad de  $P_{\theta_2\varphi_2}$ , y así sucesivamente, hasta el momento de tiempo  $n = n_\varphi$ . Finalmente, dado el modelo de aleatoriedad  $\mu$ , la probabilidad de que ocurra la sucesión de observación  $\boldsymbol{\varphi}$  se obtiene sumando la probabilidad conjunta (4.41) sobre todos las posibles sucesiones de estado  $\boldsymbol{\theta}(i)$ ,  $\forall i$ ,

$$P(\boldsymbol{\varphi}|\mu) = \sum_i P(\boldsymbol{\varphi}|\boldsymbol{\theta}(i), \mu)P(\boldsymbol{\theta}(i)|\mu) \quad (4.42)$$

La cantidad de operaciones en (4.42) es igual a  $(2n_\varphi - 1)n_\varphi^{n_\varphi}$  multiplicaciones y  $n_\varphi^{n_\varphi}$  adiciones, que hace el cálculo de la probabilidad  $P(\boldsymbol{\varphi}|\mu)$  prácticamente irrealizable. En este sentido, se emplean diferentes algoritmos recursivos que disminuyan el coste computacional.

El primer procedimiento corresponde al *algoritmo recursivo de propagación*, cuando se analiza el desarrollo de la cadena de estados en el sentido natural de desarrollo del tiempo,

y en el cual se realiza el desdoblamiento de las sucesiones de observación, empleando la relación inherente entre los elementos contiguos de las cadenas. En particular, para cada uno de los estados ocultos se puede definir la siguiente probabilidad parcial de una sucesión de observación,  $\{\varphi_1, \varphi_2, \dots, \varphi_n\}$  y del estado  $\vartheta_i$  en el momento de tiempo  $n = 1, 2, \dots$ , dado el modelo  $\mu$ ,  $\alpha_n(i) = P(\varphi_1, \varphi_2, \dots, \varphi_n, \theta_n = \vartheta_i | \mu)$ , con condición inicial para  $n = 1$ ,

$$\alpha_1(i) = p_{\theta_1}(i) P_{i\varphi_1}, \quad 1 \leq i \leq n_\vartheta$$

De tal manera, que para cada estado oculto tiene lugar la siguiente expresión recursiva

$$\alpha_{n+1}(j) = \left( \sum_{i=1}^{n_\vartheta} \alpha_n(i) \pi_{ij} \right) P_{j\varphi_{n+1}}, \quad n = 1, \dots, n_\varphi - 1, \quad 1 \leq j \leq n_\vartheta \quad (4.43)$$

La probabilidad  $P(\varphi | \mu)$  de generación de la sucesión de observación se obtiene sumando todas las probabilidades parciales con  $n = n_\varphi$  para todo el alfabeto de estados ocultos,

$$P(\varphi | \mu) = \sum_{i=1}^{n_\vartheta} \alpha_{n_\varphi}(i) \quad (4.44)$$

El algoritmo recursivo (4.43) se desarrolla en el sentido natural del tiempo y, por lo tanto, se denomina de *propagación directa*. Sin embargo, existe la versión equivalente de *propagación inversa* en el tiempo. Para esta propagación se define la probabilidad de la sucesión parcial de observación desde  $n + 1$  hasta  $n_\varphi$ , dado el estado  $\vartheta_i$  en el tiempo  $n$  y el modelo  $\mu$ ,  $\beta_n(i) = P(\varphi_{n+1}, \varphi_{n+2}, \dots, \varphi_{n_\varphi} | \theta_n = \vartheta_i, \mu)$ , con la debida condición final,  $\beta_{n_\varphi}(i) = 1$ ,  $1 \leq i \leq n_\vartheta$ .

Para cada estado oculto se tiene la siguiente expresión recursiva,

$$\beta_n(i) = \sum_{j=1}^{n_\vartheta} \pi_{ij} P_{j\varphi_{n+1}} \beta_{n+1}(j), \quad n = n_\varphi - 1, \dots, 1, \quad 1 \leq j \leq n_\vartheta \quad (4.45)$$

En cualquier caso, el algoritmo recursivo de propagación (directa o inversa) requiere tan solo de  $n_\vartheta^2 n_\varphi$  operaciones, que corresponde a un volumen aceptable de operaciones en condiciones reales de proceso.

**Selección de la mejor sucesión.** En este caso, no se tiene una solución que de forma universal, se acepte para la selección de la mejor sucesión de estados. La dificultad está en la definición de la misma sucesión óptima de estados.

Una primera solución consiste en la selección de los estados  $q_e$  que sean individualmente más probables; caso en el que se maximiza el número esperado de estados individuales correctos, para lo cual se define la variable,

$$\gamma_n(i) = P(\theta_n = \vartheta_i | \varphi, \mu) \quad (4.46)$$

esto es, la probabilidad de encontrarse en el estado  $\vartheta_i$  en el tiempo  $n$ , dada la observación

$\varphi$  y el modelo  $\mu$ . La ecuación (4.46) se expresa simplemente en términos de las variables de propagación:

$$\gamma_n(j) = \frac{\alpha_n(j)\beta_n(j)}{P(\varphi|\mu)} = \frac{\alpha_n(j)\beta_n(j)}{\sum_{i=1}^{n_\vartheta} \alpha_n(i)\beta_n(i)}$$

donde  $\alpha_n(j)$  describe la sucesión parcial de observación  $\varphi_1, \dots, \varphi_n$  y el estado  $\vartheta_j$  en tiempo  $n$ , mientras  $\beta_n(j)$  describe el resto de la sucesión de observación dado el estado  $\vartheta_j$  en  $n$ . El factor de normalización  $P(\varphi|\mu) = \sum_{i=1}^{n_\vartheta} \alpha_n(i)\beta_n(i)$ , le da el sentido de medida de probabilidad a  $\gamma_n(j)$ , de tal manera que,  $\sum_{i=1}^{n_\vartheta} \gamma_n(i) = 1$ .

Mediante el término  $\gamma_n(j)$  se puede hallar individualmente el estado  $\theta_n$  en el tiempo  $n$  más probable, como,

$$\theta_n = \arg \max_{1 \leq i \leq n_\vartheta} \{\gamma_n(i)\}, \quad 1 \leq n \leq n_\varphi \quad (4.47)$$

Aunque la expresión (4.47) maximiza el número esperado de estados correctos, (escogiendo el estado más probable por cada  $n$ ), podrían generarse algunos problemas con la sucesión de estados resultante. Por ejemplo, cuando el modelo oculto contempla transiciones de estado con valor de probabilidad cero, ( $\pi_{ij} = 0$ ), la sucesión de estados óptima ni siquiera podría ser una sucesión de estados válida. Este problema se debe a que la solución planteada en (4.47) simplemente determina el estado más probable en cada instante del tiempo, sin importar la probabilidad de ocurrencia de las sucesiones de estado.

Una posible corrección a este problema está en modificar el criterio de optimización, por ejemplo, se puede hallar la sucesión de estados que maximice el número esperado de pares de estados correctos  $(\theta_n, \theta_{n+1})$  o inclusive de cadenas más largas  $(\theta_n, \theta_{n+1}, \dots, \theta_{n+k})$ . Sin embargo, este criterio no ha mostrado ser implementable para un rango amplio de aplicaciones. Por lo tanto, en la práctica, el criterio con mayor uso se limita a hallar la mejor sucesión de estados, esto es, se maximiza  $P(\theta|\varphi, \mu)$ , considerando que es equivalente a maximizar  $P(\theta, \varphi|\mu)$ . La técnica formal de cálculo de la mejor sucesión de estados, basada en métodos de programación dinámica, corresponde al *algoritmo de Viterbi*.

Sea determinada la mejor sucesión de estados  $\theta = \{\theta_1, \dots, \theta_{n_\varphi}\}$ , para una sucesión de observación  $\varphi = \{\varphi_1, \dots, \varphi_{n_\varphi}\}$ . Se define la medida,

$$m_n(i) = \max_{\theta_1, \theta_2, \dots, \theta_{n-1}} P(\theta_1, \theta_2, \dots, \theta_n = i, \varphi_1, \varphi_2, \dots, \varphi_n | \mu)$$

que es la más alta probabilidad a lo largo de una única ruta, en el tiempo  $n$ , que da cuenta de las primeras  $n$  observaciones y finaliza en el estado  $\vartheta_i$ . De forma análoga se tiene:

$$m_{n+1}(j) = \left( \max_i m_n(i) \pi_{ij} \right) P_{j\varphi_{n+1}} \quad (4.48)$$

La reconstrucción de la sucesión de estados exige que se conserve el argumento que maximiza (4.48), para cada  $n$  y  $j$  lo que se puede implementar mediante un arreglo  $\psi_n(j)$ ,

calculado de forma recursiva como

$$\psi_n(j) = \arg \max_i (m_{n-1}(i)\pi_{ij}), \quad n = 2, \dots, n_\varphi - 1, \quad 1 \leq j \leq n_\vartheta$$

con valores iniciales,  $\psi_1(j) = 0$ ,  $1 \leq j \leq n_\vartheta$ .

**Ajuste de los parámetros del modelo.** Se considera que no existe una solución analítica para un modelo que maximice la probabilidad de la sucesión de observación. Esto es, dada cualquier sucesión de observación sobre un intervalo finito de tiempo, durante el entrenamiento del sistema, no hay una forma óptima para la estimación de los parámetros del modelo. Sin embargo, se puede determinar el modelo  $\mu = \{\Pi, \mathbf{P}(\mathbf{v}|\boldsymbol{\theta}), \mathbf{p}_{\theta_1}\}$ , tal que se maximiza localmente la función  $P(\varphi|\mu)$ , usando diferentes procedimientos iterativos de estimación existentes.

El primer método corresponde al método de *máxima esperanza* que implica la estimación de máxima verosimilitud y se emplea cuando la información necesaria para estimar los parámetros del modelo está incompleta, como es el caso de los modelos ocultos de Markov, en el que se conocen las sucesiones de observación, pero no se conocen las sucesiones de estados. El algoritmo se compone de dos pasos:

- *Cálculo de la esperanza* o promedio,  $\mathcal{Q}(\tilde{\mu}|\mu) = E\{\log P(\boldsymbol{\theta}, \varphi|\tilde{\mu})|\varphi, \mu\}$ ,  $\mu, \tilde{\mu} \in \mathfrak{M}$
- *Maximización*, cuando se escogen los valores de  $\tilde{\mu}$  que maximicen un funcional  $\mathcal{Q}(\tilde{\mu}|\mu)$ .

donde  $\tilde{\mu}$  representa el nuevo conjunto de parámetros después de una iteración y  $\mathfrak{M}$  es el espacio de los parámetros del modelo. El algoritmo de máxima esperanza se basa en la búsqueda de un conjunto de parámetros  $\tilde{\mu}$  que maximice  $\log P(\boldsymbol{\theta}, \varphi|\tilde{\mu})$ . Sin embargo, debido a que no se conoce la relación  $\log P(\boldsymbol{\theta}, \varphi|\tilde{\mu})$ , entonces se maximiza su valor esperado actual dados, tanto las observaciones  $\varphi$  como el modelo  $\mu$ . La función  $\mathcal{Q}(\tilde{\mu}|\mu)$  se conoce como la *función auxiliar*.

Con el fin de describir el procedimiento de estimación recursiva de los parámetros del modelo oculto de Markov, se define  $\zeta_n(i, j)$  como la probabilidad de situarse en el estado  $\vartheta_i$  en el tiempo  $n$  y en el estado  $\vartheta_j$  para el tiempo  $n + 1$ , dado el modelo y la sucesión de observación

$$\zeta_n(i, j) = P(\theta_n = \vartheta_i, \theta_{n+1} = \vartheta_j | \varphi, \tilde{\mu})$$

la cual puede ser descrita en términos de las variables de propagación en la forma,

$$\zeta_n(i, j) = \frac{\alpha_n(i)\pi_{ij}p_{j\varphi_{n+1}}\beta_{n+1}(j)}{P(\varphi|\tilde{\mu})} = \frac{\alpha_n(i)\pi_{ij}p_{j\varphi_{n+1}}\beta_{n+1}(j)}{\sum_{i=1}^{n_\vartheta} \sum_{j=1}^{n_\vartheta} \alpha_n(i)\pi_{ij}p_{j\varphi_{n+1}}\beta_{n+1}(j)}$$

donde el numerador corresponde a  $P(\theta_n = \vartheta_i, \theta_{n+1} = \vartheta_j | \varphi, \tilde{\mu})$  que dividido por  $P(\varphi|\tilde{\mu})$  es

la medida deseada de probabilidad, la cual, se relaciona con la medida (4.46),

$$\gamma_n(i) = \sum_{j=1}^{n_\vartheta} \zeta_n(i, j)$$

La probabilidad conjunta de la sucesión de observación  $\varphi$  y la sucesión de estados  $\theta$  bajo los parámetros del modelo  $\tilde{\mu}$  está dada por la ecuación (4.41) que toma la forma

$$P(\varphi, \theta | \tilde{\mu}) = p_{\theta_1}(\theta_1) \prod_{n=2}^{n_\varphi} P(\theta_{n+1} | \theta_n, \tilde{\mu}) \prod_{n=2}^{n_\varphi} P(\varphi_n | \theta_n, \tilde{\mu})$$

El logaritmo de la función de probabilidad conjunta está dado por

$$\log P(\varphi, \theta | \tilde{\mu}) = \log p_{\theta_1}(\theta_1) + \sum_{n=2}^{n_\varphi} \log P(\theta_{n+1} | \theta_n, \tilde{\mu}) + \sum_{n=2}^{n_\varphi} \log P(\varphi_n | \theta_n, \tilde{\mu}).$$

El algoritmo de máxima esperanza requiere la maximización de la función  $\mathcal{Q}(\tilde{\mu} | \mu)$ , que es el valor esperado del logaritmo de la probabilidad conjunta, donde el valor esperado se toma con respecto a la distribución anterior de los estados,  $P(\theta | \varphi, \mu)$ , dadas las observaciones. Así,

$$\begin{aligned} \mathcal{Q}(\tilde{\mu} | \mu) = \sum_{\theta} P(\theta | \varphi, \mu) \log \tilde{p}_{\theta_1}(\theta_1) + \sum_{\theta} P(\theta | \varphi, \mu) \sum_{n=2}^{N_\varphi} \log P(\theta_n | \theta_{n-1}, \tilde{\mu}) + \\ + \sum_{\theta} P(\theta | \varphi, \mu) \sum_{n=1}^{N_\varphi} \log P(\varphi_n | \theta_n, \tilde{\mu}) \end{aligned} \quad (4.49)$$

donde  $\sum_{\theta}$  denota la suma sobre todas las posibles sucesiones de estados. La ecuación (4.49) está compuesta de tres términos; el primero para los parámetros de la probabilidad inicial  $\mathbf{p}_{\theta_1}$ , el segundo para los parámetros de la matriz probabilidad de transición de estados  $\mathbf{\Pi}$  y el tercero para la matriz de probabilidad del modelo de observación  $\mathbf{P}(\mathbf{v} | \boldsymbol{\vartheta})$ ,

$$\mathcal{Q}(\tilde{\mu} | \mu) = \mathcal{Q}(\tilde{\mathbf{p}}_{\theta_1} | \mathbf{p}_{\theta_1}) + \mathcal{Q}(\tilde{\mathbf{\Pi}} | \mathbf{\Pi}) + \mathcal{Q}(\tilde{\mathbf{P}}(\mathbf{v} | \boldsymbol{\vartheta}) | \mathbf{P}(\mathbf{v} | \boldsymbol{\vartheta})) \quad (4.50)$$

Los términos en la ecuación (4.50) se pueden maximizar de forma separada. La expresión de la estimación recursiva para la densidad inicial de probabilidades  $\mathbf{p}_{\theta_1}$  puede obtenerse como

$$\mathcal{Q}(\tilde{\mathbf{p}}_{\theta_1} | \mathbf{p}_{\theta_1}) = \sum_{\theta} P(\theta | \varphi, \mu) \log \tilde{p}_{\theta_1}(\theta_1)$$

que es equivalente a la expresión

$$\mathcal{Q}(\tilde{\mathbf{p}}_{\theta_1} | \mathbf{p}_{\theta_1}) = \sum_{i=1}^{n_\vartheta} P(\theta_1 = \vartheta_i | \varphi, \mu) \log \tilde{p}_{\theta_1}(\theta_1 = \vartheta_i)$$

que, usando la ecuación (4.46), es igual a

$$\mathcal{Q}(\tilde{\mathbf{p}}_{\theta_1} | \mathbf{p}_{\theta_1}) = \sum_{i=1}^{n_{\vartheta}} \gamma_1(i) \log \tilde{p}_{\theta_1}(i)$$

La suma del factor de Lagrange  $\varrho$  en la expresión anterior, asumiendo la debida restricción,  $\sum_i p_{\theta_1}(i) = 1$  y al hacer la derivada igual a cero, resulta en

$$\frac{\partial}{\partial \tilde{p}_{\theta_1}(i)} \left( \sum_{i=1}^{n_{\vartheta}} \gamma_1(i) \log \tilde{p}_{\theta_1}(i) + \varrho \left( \sum_i \tilde{p}_{\theta_1}(i) - 1 \right) \right) = 0$$

Del desarrollo de la derivada, realizando la suma sobre  $i$  para obtener  $\varrho$  y resolviendo para  $\tilde{p}_1^{\theta}(i)$ , se obtiene la solución

$$\tilde{p}_{\theta_1}(i) = \gamma_1(i) \quad (4.51)$$

que puede calcularse usando las variables de propagación  $\alpha_1(i)$  y  $\beta_1(i)$ . En el procedimiento de promedio del algoritmo se calcula  $\gamma_1(i)$  usando el conjunto de parámetros  $\mu$ . Mientras, en el procedimiento de maximización se basa en (4.51) para obtener  $\tilde{\mu}$ .

La expresión para la actualización de los parámetros de la matriz de transición de estados  $\tilde{\Pi}$  puede encontrarse al maximizar la función,

$$\mathcal{Q}(\tilde{\Pi} | \Pi) = \sum_{\theta} P(\theta | \varphi, \mu) \sum_{n=2}^{N_{\varphi}} \log P(\theta_n | \theta_{n-1}, \tilde{\mu})$$

Sin embargo, la anterior suma puede descomponerse en la forma:

$$\begin{aligned} \mathcal{Q}(\tilde{\Pi} | \Pi) &= \sum_{i=1}^{n_{\vartheta}} \sum_{j=1}^{n_{\vartheta}} \sum_{n=2}^{n_{\varphi}} P(\theta_n = \vartheta_j, \theta_{n-1} = \vartheta_i | \mu) \log P(\theta_n | \theta_{n-1}, \tilde{\mu}) \\ &= \sum_{i=1}^{n_{\vartheta}} \sum_{j=1}^{n_{\vartheta}} \sum_{n=2}^{n_{\varphi}} \zeta_n(i, j) \log(\tilde{\pi}_{ij}) \end{aligned}$$

entonces, la maximización de la expresión anterior, con la restricción  $\sum_{j=1}^{n_{\vartheta}} \pi_{ij} = 1$ , conduce a la siguiente forma de actualización

$$\tilde{\pi}_{ij} = \frac{\sum_{n=2}^{n_{\varphi}} \zeta_n(i, j)}{\sum_{n=2}^{n_{\varphi}} \gamma_n(i)} \quad (4.52)$$

En este caso, durante la etapa de promedio, los valores de  $\zeta_n(i, j)$  y  $\gamma_n(i, j)$  se estiman usando los parámetros del modelo  $\mu$ . Mientras, para la maximización, se emplea la expresión (4.52) para calcular los nuevos parámetros  $\tilde{\mu}$ .

De otra parte, la expresión para la actualización de los parámetros del modelo de obser-

vación  $\tilde{\mathbf{P}}(\mathbf{v}|\boldsymbol{\theta})$  puede encontrarse maximizando

$$\mathcal{Q}(\tilde{\mathbf{P}}(\mathbf{v}|\boldsymbol{\theta})|\mathbf{P}(\mathbf{v}|\boldsymbol{\theta})) = \sum_{\boldsymbol{\theta}} P(\boldsymbol{\theta}|\varphi, \mu) \sum_{n=1}^{n_{\varphi}} \log P(\varphi_n|\theta_n, \tilde{\mu})$$

De manera similar, se descompone la anterior suma,

$$\begin{aligned} \mathcal{Q}(\tilde{\mathbf{P}}(\mathbf{v}|\boldsymbol{\theta})|\mathbf{P}(\mathbf{v}|\boldsymbol{\theta})) &= \sum_{i=1}^{n_{\vartheta}} \sum_{n=1}^{n_{\varphi}} P(\theta_n = \vartheta_i|\mu) \log P(\varphi_n = v_k|\theta_n = \vartheta_i, \tilde{\mu}) \\ &= \sum_{i=1}^{n_{\vartheta}} \sum_{n=1}^{n_{\varphi}} \gamma_n(i) \log \tilde{p}_{ik} \end{aligned}$$

con lo cual, la maximización de la expresión anterior, dada la restricción

$$\sum_{k=1}^{n_{\varphi}} \tilde{P}_{ik} = 1$$

resulta en:

$$\tilde{P}_{ik} = \frac{\sum_{n=1}^{n_{\varphi}} \gamma_n(i) \delta(\varphi_n, v_k)}{\sum_{n=1}^{n_{\varphi}} \gamma_n(i)}, \quad \delta(\varphi_n, v_k) = \begin{cases} 1, & \varphi_n = v_k \\ 0, & \varphi_n \neq v_k \end{cases} \quad (4.53)$$

En este caso, la función  $\Lambda_{\varphi}(\mu) = \log P(\varphi|\mu)$  corresponde a la respectiva relación de verosimilitud,

$$\Lambda_{\varphi}(\mu) = \mathcal{Q}(\tilde{\mu}|\mu) - \mathcal{H}(\tilde{\mu}|\mu),$$

donde  $\mathcal{H}(\tilde{\mu}|\mu) = E\{\log P(\boldsymbol{\theta}|\varphi, \tilde{\mu})|\varphi, \mu\}$ .

A partir de la desigualdad de Jensen [34] se obtiene  $\mathcal{H}(\tilde{\mu}|\mu) \leq \mathcal{H}(\mu|\mu)$  para cualquier  $\mu$  y  $\tilde{\mu}$  en  $\mathfrak{M}$ , cumpliéndose la igualdad, si y sólo si,  $P(\boldsymbol{\theta}|\varphi, \mu) = P(\boldsymbol{\theta}|\varphi, \tilde{\mu})$ . En efecto, para cualquier  $\tilde{\mu}$ ,

$$\begin{aligned} \mathcal{H}(\tilde{\mu}|\mu) - \mathcal{H}(\mu|\mu) &= E \left\{ \log \left( \frac{P(\boldsymbol{\theta}|\varphi, \tilde{\mu})}{P(\boldsymbol{\theta}|\varphi, \mu)} \right) \middle| \varphi, \mu \right\} \\ &\leq \log E \left\{ \left( \frac{P(\boldsymbol{\theta}|\varphi, \tilde{\mu})}{P(\boldsymbol{\theta}|\varphi, \mu)} \right) \middle| \varphi, \mu \right\} \\ &= \log \int_{\boldsymbol{\theta}} \left( \frac{P(\boldsymbol{\theta}|\varphi, \tilde{\mu})}{P(\boldsymbol{\theta}|\varphi, \mu)} \right) P(\boldsymbol{\theta}|\varphi, \mu) d\boldsymbol{\theta} = \log \int_{\boldsymbol{\theta}} P(\boldsymbol{\theta}|\varphi, \tilde{\mu}) d\boldsymbol{\theta} \\ &= 0, \end{aligned}$$

usando igualmente el hecho de la concavidad de la función logarítmica.

En los modelos en que se desarrolla el método de máxima esperanza, se asume la posi-

bilidad de evaluación de las derivadas de la relación  $\Lambda_\varphi(\mu)$  con respecto a los parámetros del modelo  $\mu$ . Como consecuencia de lo anterior, en lugar de recurrir a cualquier algoritmo específico, por ejemplo, el de máxima esperanza, directamente se pueden utilizar métodos de optimización basados en el análisis del gradiente. Concretamente, el algoritmo simple de gradiente descendente, aunque no sea el más eficiente. Sin embargo, en este método, el valor presente de los parámetros  $\tilde{\mu}_n$  se actualiza sumándole un valor de ajuste que multiplica al gradiente, en la siguiente forma recursiva:

$$\tilde{\mu}_{n+1} = \tilde{\mu}_n + \rho_n \nabla_\mu \Lambda_\varphi(\tilde{\mu}_n), \quad \rho_n > 0$$

El múltiplo  $\rho_n$  es un escalar no negativo que necesita ajustarse en cada iteración para asegurar que la sucesión  $\{\Lambda_\varphi(\tilde{\mu}_n)\}$  no sea decreciente. De cualquier manera, el algoritmo de gradiente descendente suele ser poco utilizado debido a su lenta convergencia. Para garantizar valores más rápidos de convergencias se han desarrollado diferentes algoritmos basados en propiedades de segundo orden de la función objetivo, por ejemplo, el algoritmo de Newton o el de Newton-Raphson,

$$\tilde{\mu}_{n+1} = \tilde{\mu}_n - H^{-1}(\tilde{\mu}_n) \nabla_\mu \Lambda_\varphi(\tilde{\mu}_n),$$

donde  $H^{-1}(\tilde{\mu}_n) = \nabla_\mu^2 \Lambda_\varphi(\tilde{\mu}_n)$  es la matriz Hessiana de la función objetivo. El algoritmo iterativo de Newton está basada en la siguiente aproximación de segundo orden:

$$\Lambda(\mu) \approx \Lambda(\mu') + \nabla \Lambda(\mu')(\mu - \mu') + \frac{1}{2}(\mu - \mu')^\top H(\mu')(\mu - \mu'),$$

Si la sucesión  $\{\tilde{\mu}_n\}_{n>0}$  producida por el algoritmo converge a un punto  $\tilde{\mu}^*$ , en el cual la matriz Hessiana es negativa definida, la convergencia es, por lo menos cuadrática. Por lo anterior, el procedimiento puede ser bastante eficiente. Sin embargo, una restricción seria del algoritmo de Newton se encuentra en el hecho que, excepto en el caso particular en que la función  $\Lambda(\mu)$  sea estrictamente cóncava, pueden existir regiones del espacio paramétrico en las cuales la matriz Hessiana no tenga inversa o no sea negativa semi-definida, lo que conduce a inestabilidades numéricas. Para solucionar este problema, los métodos cuasi-Newton introducen la recursión modificada

$$\tilde{\mu}_{n+1} = \tilde{\mu}_n + \rho_n W^n(\tilde{\mu}_n) \nabla_\mu \Lambda_\varphi(\tilde{\mu}_n),$$

donde  $W^n$  es una matriz de pesos que puede calcularse en cada iteración, igual que  $\rho_n$ , un factor que se usa para calcular la longitud del paso en la dirección de búsqueda. Si  $W^n$  se encuentra cercano a  $-H^{-1}(\tilde{\mu}_n)$  cuando ocurre la convergencia, el algoritmo modificado compartirá las mismas propiedades deseables del algoritmo de Newton. De otra parte, usando una matriz  $W^n$  diferente de  $-H^{-1}(\tilde{\mu}_n)$ , las restricciones numéricas asociadas con la inversión de la matriz se pueden evitar. Estos métodos construyen  $W^n$  a partir de la información del gradiente, sin la evaluación directa de la matriz Hessiana.



**Ejercicio en el CP 4.2.** Los ejemplos y ejercicios de este capítulo se realizan con base en el *toolbox*, especialmente desarrollado en MATLAB para el estudio de los modelos ocultos de Markov, que puede ser descargado en: <http://www.cs.ubc.ca/~murphyk/Software/HMM/hmm.html>. El primer paso en el entrenamiento de los modelos ocultos de Markov es su inicialización, para lo cual es necesario definir el número de estados del modelo y el número de componentes Gaussianas por estado.

```
% Parámetros del modelo de Markov:
M = 3;           % Número de componentes Gaussianas por estado
Q = 2;           % Número de estados
cov_type = 'full'; % Tipo de matriz de covarianza
```

Además, es necesario definir el tipo de matriz de covarianza que será utilizada en las componentes Gaussianas, que puede ser de los tipos: completa ('full'), diagonal ('diag') y esférica ('spherical'). A modo de ejemplo, se escoge el tipo de matriz de covarianza ('full') por cuanto permite modelar la relación entre cada una de las variables. Sin embargo, en la práctica, utilizar una matriz de covarianza completa, impide el buen desempeño del algoritmo de estimación presentándose problemas de singularidad. Por lo tanto, en muchos problemas reales es más común emplear el tipo de matriz de covarianza diagonal, en la cual se asume independencia entre los parámetros. Así mismo, se asume que se tiene un conjunto de entrenamiento y que cada trayectoria está representada por una sucesión de observaciones. La longitud de cada sucesión de observaciones se representa por T. Mientras, cada observación está compuesta por O coeficientes o parámetros. Como ejemplo concreto se generan Ns sucesiones de observaciones con parámetros:

```
% Conjunto de entrenamiento
Ns = 50; T = 10; O = 8; data = randn(O,T,Ns);
prior0 = normalise(rand(Q,1)); % vector probabilidad inicial
transmat0 = mk_stochastic(rand(Q,Q)); % Matriz de probabilidades de
                                         % transición
[mu0, Sigma0] = mixgauss_init(Q*M, data, cov_type); % media y covarianza
```

La inicialización de los parámetros del modelo (el vector probabilidad inicial, la matriz de probabilidades de transición y los parámetros de las componentes Gaussianas: vector de pesos, vector de medias y matriz de covarianza, por cada uno de los estados) debe tener en cuenta las restricciones que conservan las medidas estadísticas.

```
% Restricciones de las medidas estadísticas
mu0 = reshape(mu0, [O Q M]); Sigma0 = reshape(Sigma0, [O O Q M]);
mixmat0 = mk_stochastic(rand(Q,M)); % Vector de pesos de las mezclas Gaussianas
                                         % Entrenamiento
[LL, prior, transmat, mu, Sigma, mixmat] = ...
    mhmm_em(data, prior0, transmat0, mu0, Sigma0, mixmat0, 'max_iter',5);
% Evaluación
loglik = mhmm_logprob(data, prior, transmat, mu, Sigma, mixmat);
```

El parámetro LL corresponde al logaritmo de la verosimilitud en el entrenamiento. La evaluación de la pertenencia de una sucesión al modelo se realiza mediante el cálculo de la máxima probabilidad a posteriori de la sucesión, dado un modelo concreto.

### 4.2.3. Clasificación y comparación modelos

Hasta ahora, se han considerado las cadenas ocultas de Markov ergódicas, en las cuales a cada estado de la cadena se puede llegar en un único paso desde cualquier otro estado del modelo (estrictamente hablando, un modelo ergódico tiene la propiedad por la cual cada estado puede alcanzarse desde cualquier otro estado después de un número finito de pasos).

Así mismo, los modelos de cambio de estado también se orientan en el sentido de los valores del proceso (modelos de nacimiento y muerte). En la práctica, mayor uso han encontrado los modelos de nacimiento, cuando los coeficientes de transición de estado tienen la propiedad,  $\pi_{ij} = 0$ ,  $j < i$ , esto es, no se permiten cambios a estados con valores del proceso menores que el actual. Entonces, el vector de probabilidad de estado inicial es:

$$p_{\theta_1}(i) = \begin{cases} 0, & i \neq 1 \\ 1, & i = 1 \end{cases}$$

que implica que el proceso debe comenzar desde el estado 1 y a finalizar en el estado  $n_g$ .

Sobre los modelos de nacimiento, se suelen imponer restricciones adicionales para asegurar que no ocurran cambios mayores entre los índices de los estados. En este sentido, es usual el empleo de la restricción de la forma  $\pi_{ij} = 0$ ,  $j > i + \Delta$ . En el caso particular de los modelos de nacimiento con  $\Delta = k$ , se obtiene la siguiente matriz de transición de estado:

$$\begin{bmatrix} \pi_{11} & \pi_{12} & \cdots & \pi_{1k} & 0 & \cdots & 0 \\ 0 & \pi_{22} & \cdots & \pi_{2k} & \pi_{2(k+1)} & \cdots & 0 \\ \vdots & \vdots & & \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & 0 & 0 & \cdots & \pi_{kk} \end{bmatrix}$$

Las anteriores restricciones, impuestas sobre la orientación de la forma de cambio de los valores del modelo no afectan, en general, el procedimiento de estimación de sus parámetros.

**Densidades continuas de observación.** Las observaciones, de las cuales se estiman los parámetros de los modelos ocultos de Markov, pueden corresponder a eventos discretos, cuando se tiene un alfabeto finito a priori definido de representación del proceso y, por lo tanto, existe una FDP discreta para cada estado del modelo. En la práctica, es frecuente que las observaciones correspondan a procesos aleatorios continuos. La primera forma de análisis de estos procesos consiste en la construcción de un alfabeto via la cuantificación del proceso, cuya principal restricción está en la pérdida de información asociada a la aproximación. En este sentido, es preferible la descripción de los procesos aleatorios en los modelos ocultos de Markov mediante FDP continuas, teniendo en cuenta las restricciones necesarias para lograr que los procedimientos de estimación recursiva sean consistentes. La forma más común de representación de la FDP corresponde al modelo de mezclas finitas:

$$p_{i\varphi_n} = \sum_{m=1}^M c_{jm} \mathcal{N}_{\varphi_n}(m_{jm}, \mathbf{K}_{jm})$$

donde  $\varphi_n$  es el vector que se está modelando en el instante  $n$ ,  $c_{jm}$  es el coeficiente de mezcla para la  $m$  componente en el estado  $j$  y  $\mathcal{N}$  es cualquier densidad elípticamente simétrica o log-cóncava, con media  $\mathbf{m}_{jm}$  y matriz de covarianza  $\mathbf{K}_{jm}$  para la componente  $m$  de la mezcla en el estado  $j$ . Usualmente, en calidad de densidad  $\mathcal{N}$  se emplea la Gaussiana.

Los coeficientes de la mezcla  $c_{jm}$  deben satisfacer la siguiente restricción estocástica:  $\sum_{m=1}^M c_{jm} = 1$ ,  $1 \leq j \leq n_\vartheta$ , con  $c_{jm} \geq 0$ ,  $1 \leq j \leq n_\vartheta$  y  $1 \leq m \leq M$ , de tal manera, que la FDP sea adecuadamente normalizada:  $\int_{-\infty}^{\infty} \mathbf{p}_i \varphi d\varphi = 1$ .

Las relaciones de estimación recursiva para los coeficientes de las densidades de mezcla,  $c_{jm}$ ,  $\mathbf{m}_{jm}$ , y  $\mathbf{K}_{jk}$ , para el caso de FDP Gaussianas, empleando el algoritmo de máxima esperanza, se demuestran ser iguales a

$$\tilde{c}_{jk} = \frac{\sum_{n=1}^{n_\varphi} \gamma_n(j, k)}{\sum_{n=1}^{n_\varphi} \sum_{k=1}^M \gamma_n(j, k)} \quad (4.54a)$$

$$\tilde{\mathbf{m}}_{jk} = \frac{\sum_{n=1}^{n_\varphi} \gamma_n(j, k) \varphi_n}{\sum_{n=1}^{n_\varphi} \gamma_n(j, k)} \quad (4.54b)$$

$$\tilde{\mathbf{K}}_{jk} = \frac{\sum_{n=1}^{n_\varphi} \gamma_n(j, k) (\varphi_n - \mathbf{m}_{jk})(\varphi_n - \mathbf{m}_{jk})^\top}{\sum_{n=1}^{n_\varphi} \gamma_n(j, k)} \quad (4.54c)$$

donde  $\gamma_n(j, k)$  es la probabilidad de la observación  $\varphi_n$  dada por la componente de mezcla  $k$  del estado  $j$ , esto es:

$$\gamma_n(j, k) = \left( \frac{\alpha_n(j) \beta_n(j)}{\sum_{i=1}^{n_\vartheta} \alpha_n(i) \beta_n(i)} \right) \left( \frac{c_{jk} \mathcal{N}_{\varphi_n}(\mathbf{m}_{jk}, \mathbf{K}_{jk})}{\sum_{m=1}^M c_{jm} \mathcal{N}_{\varphi_n}(\mathbf{m}_{jm}, \mathbf{K}_{jm})} \right) \quad (4.55)$$

El término  $\gamma_n(j, k)$  generaliza la variable  $\gamma_n(i)$  en (4.46) que es el caso de una mezcla simple (un solo componente) o de una FDP discreta. La estimación de los valores  $\pi_{ij}$  se mantiene igual que para la FDP discreta de observación, (4.52). En el algoritmo de máxima esperanza, para la estimación del promedio  $\tilde{\mu}$ , se calcula  $\gamma_n(j, k)$  mediante los parámetros  $\mu$ , mientras en la maximización se usan las ecuaciones (4.54a), (4.54b) y (4.54c).

La estimación recursiva de  $c_{jk}$  es la relación entre el número esperado de veces que el sistema está en el estado  $j$  empleando la componente de mezcla  $k$ , y el número esperado de veces que está en el estado  $j$ . La estimación recursiva para el vector medio  $\mathbf{m}_{jk}$  pondera cada término del numerador de (4.54b) por cada observación, por lo que resulta un valor esperado de la porción del vector de observación que describe la componente de mezcla  $k$ . Así mismo, ocurre la interpretación para la matriz de covarianza  $\mathbf{K}_{jk}$  obtenida.

La presentación usual de las expresiones de actualización se realiza en términos de los vectores de observación  $\varphi_n$  (ecs. (4.54a), (4.54b), (4.54c)). Sin embargo, es posible expresar  $\mathbf{m}_{jk}$  y  $\mathbf{K}_{jk}$  en términos de *estadísticas suficientes promedio*, cuya estimación es independiente del número de observaciones,  $n_\varphi$ :

$$\tilde{\mathbf{m}}_{jk} = \frac{s_{\varphi,jk}}{\gamma(j,k)}, \quad (4.56a)$$

$$\tilde{\mathbf{K}}_{jk} = \frac{s_{\varphi\varphi^\top,jk}}{\gamma(j,k)} - \tilde{\mathbf{m}}_{jk}\tilde{\mathbf{m}}_{jk}^\top, \quad (4.56b)$$

donde las estadísticas suficientes se dan de la forma:

$$\gamma(j,k) = \sum_{n=1}^{n_\varphi} \gamma_n(j,k), \quad s_{\varphi,jk} = \sum_{n=1}^{n_\varphi} \gamma_n(j,k)\varphi_n, \quad s_{\varphi\varphi^\top,jk} = \sum_{n=1}^{n_\varphi} \gamma_n(j,k)\varphi_n\varphi_n^\top$$

En la estimación del promedio, el algoritmo de máxima esperanza calcula los valores  $\gamma_n(j,k)$ , las estadísticas suficientes  $\gamma(j,k)$ ,  $s_{\varphi,jk}$  y  $s_{\varphi\varphi^\top,jk}$ . Mientras, el procedimiento de maximización se basa en (4.56a) y (4.56b) para obtener  $\tilde{\boldsymbol{\mu}}$ . Desde el punto de vista de coste computacional, el reemplazo de (4.54b), (4.54c) para la maximización, resulta en un cálculo más rápido de los parámetros debido a que no se hace la suma  $\sum_n$ , por cuanto esa información está ya contenida en las estadísticas suficientes. Así, el número de operaciones en la maximización permanece constante y no depende del número de observaciones  $n_\varphi$ .

**Modelos ocultos de Markov autorregresivos.** Sea el vector de observación  $\varphi$  con componentes  $(x_0, x_1, \dots, x_{K-1})$  generado sobre un espacio con dimensión  $K$ . La FDP base para el vector de observación se considera del tipo Gaussiana, y se asume el modelo autorregresivo de orden  $p$  que relaciona los componentes de  $\varphi$  en la forma:

$$x_k = - \sum_{i=1}^p a_i x_{k-i} + e_k \quad (4.57)$$

donde  $e_k$ ,  $k = 0, \dots, K-1$  son variables aleatorias independientes Gaussianas, con media 0 y varianza  $\sigma_e^2$ ,  $a_i$ ,  $i = 1, \dots, p$  son los coeficientes de predicción o autorregresión. La FDP de  $\varphi$ , para un valor grande de  $K$ , se aproxima como,  $p(\varphi) = (2\pi\sigma_e^2)^{-K/2} \exp\{-\frac{1}{2\sigma_e^2}r(\varphi, \mathbf{a})\}$ , donde

$$r(\varphi, \mathbf{a}) = R_a(0)R(0) + 2 \sum_{i=1}^p R_a(i)R(i), \quad \mathbf{a} = [1, a_1, a_2, \dots, a_p]^\top, \quad (4.58a)$$

$$R_a(i) = \sum_{j=0}^{p-i} a_j a_{j+i}, \quad 1 \leq i \leq p \quad (4.58b)$$

$$R(i) = \sum_{j=0}^{k-i-1} x_j x_{j+i}, \quad 0 \leq i \leq p \quad (4.58c)$$

En las anteriores expresiones,  $R(i)$  es la función de correlación propia de las observaciones y  $R_a(i)$  es la función de correlación propia de los coeficientes autorregresivos. El valor residual total de predicción por ventana del proceso se determina como

$$\alpha_r = E\left\{\sum_{i=1}^k (e_i)^2\right\} = k\sigma_e^2$$

donde  $\sigma_e^2$  es la varianza por cada observación de la señal de error. Asumiendo vectores normalizados de observación,  $\tilde{\varphi} = \varphi/\sqrt{\alpha_r} = \varphi/\sqrt{k\sigma_e^2}$ , entonces la FDP toma la forma:

$$p(\tilde{\varphi}) = \left(\frac{2\pi}{k}\right)^{-k/2} \exp\left\{-\frac{k}{2}r(\tilde{\varphi}, \mathbf{a})\right\} \quad (4.59)$$

En la práctica, el factor  $k$  en (4.59) se reemplaza por la longitud efectiva de la ventana  $\tilde{k}$  que corresponde a la longitud efectiva de cada vector de observación. En el análisis de los modelos autoregresivos ocultos de Markov se asume una FDP de mezcla en la forma:

$$p_{i\varphi} = \sum_{m=1}^M c_{jm} p_{\varphi}(i; m)$$

donde cada término  $p_{\varphi}(i; m)$  es la FDP definida en (4.59) con su respectivo vector de autorregresión  $\mathbf{a}_{jm}$  o bien de forma equivalente por el vector de correlación propia,  $\mathbf{R}_{jm}$ :

$$p_{\varphi}(i; m) = \left(\frac{2\pi}{k}\right)^{-k/2} \exp\left\{-\frac{k}{2}r(\varphi, \mathbf{a}_{jm})\right\}$$

La expresión de estimación recursiva para la sucesión con correlación propia,  $R(i)$  en (4.58a), dados el estado  $j$  y la mezcla  $k$ , tiene la forma:

$$\tilde{\mathbf{R}}_{jk} = \frac{\sum_{n=1}^{N_{\varphi}} \gamma_n(j, k) \mathbf{R}_n}{\sum_{n=1}^{N_{\varphi}} \gamma_n(j, k)}$$

donde  $\mathbf{R}_n = [R_n(0), R_n(1), \dots, R_n(p)]^T$  es el vector de correlación propia definido por (4.58c) para la ventana  $n$ , mientras  $\gamma_n(j, k)$  es definida como la probabilidad de encontrarse en el estado  $j$  en el tiempo  $n$  y usando la componente de mezcla  $k$ , esto es

$$\gamma_n(j, k) = \left(\frac{\alpha_n(j)\beta_n(j)}{\sum_{j=1}^{n_{\Theta}} \alpha_n(j)\beta_n(j)}\right) \left(\frac{c_{jk}p_{\varphi_n}(j; k)}{\sum_{k=1}^M c_{jk}p_{\varphi_n}(j; k)}\right) \quad (4.60)$$

De (4.60) se observa que  $\tilde{\mathbf{R}}_{jk}$  es una suma ponderada (por la probabilidad de ocurrencia) de las funciones de correlación propia normalizadas de las ventanas en la sucesión de

observación. A partir de  $\tilde{\mathbf{R}}_{jk}$ , se resuelve el conjunto de ecuaciones normales para obtener el vector de coeficientes autorregresivos  $\tilde{\mathbf{a}}_{jk}$ , dada la  $k$  mezcla del estado  $j$ , los cuales se calculan empleando (4.58b), cerrando así el bucle de estimación.

**Estimación óptima de los parámetros del modelo.** La efectividad de los modelos ocultos de Markov está relacionada con dos aspectos básicos: el mismo proceso estocástico que cumple las restricciones del modelo y la estimación confiable de los parámetros del modelo. En este sentido, se han desarrollado diversos métodos de estimación óptima alternativos al procedimiento básico de máxima verosimilitud.

La primera aproximación consiste en generar de forma conjunta varios modelos  $\mu_\nu$ ,  $\nu = 1, \dots, V$ , de manera tal que se maximice su poder discriminante, es decir, la habilidad de cada modelo para distinguir entre sucesiones de observación generadas por el modelo correcto y aquellas generadas por modelos alternativos. El criterio básico de máxima verosimilitud consiste en el empleo de sucesiones de observación separadas  $\varphi^\nu$  a fin de derivar los parámetros de modelo para cada modelo  $\mu_\nu$ . Este criterio de optimización conduce a

$$P_\nu^* = \max_{\mu_\nu} P(\varphi^\nu | \mu_\nu)$$

Un criterio de optimización alternativo es el criterio de máximo de información mutua en el cual la información mutua promedio  $I$  entre la sucesión de observación  $\varphi^\nu$  y el conjunto completo de modelos  $\mu_C = (\mu_1, \mu_2, \dots, \mu_V)$  se maximiza.

La implementación del criterio de máximo de información mutua, en concordancia con la definición (2.57), es la medida:

$$I(\varphi^\nu; \mu_w) = \max_{\mu_C} \left\{ P(\varphi^\nu | \mu_\nu) - \log \sum_{w=1}^V P(\varphi^\nu | \mu_w) \right\} \quad (4.61)$$

esto es, escoger  $\mu_C$  para separar el modelo correcto  $\mu_\nu$  de todos los demás modelos alternativos sobre la sucesión de entrenamiento  $\varphi^\nu$ . Sumando (4.61) sobre todas las sucesiones de entrenamiento, se puede encontrar el conjunto de modelos con mayor separación,

$$I(\varphi^\nu; \mu_w) = \max_{\mu_C} \left\{ \sum_{\nu=1}^V \left( P(\varphi^\nu | \mu_\nu) - \log \sum_{w=1}^V P(\varphi^\nu | \mu_w) \right) \right\} \quad (4.62)$$

Hay diversas razones teóricas por las cuales no existen soluciones analíticas ni recursivas de estimación, para obtener los parámetros mediante el criterio de máximo de información mutua. En la práctica, la solución de (4.62) se lleva a cabo mediante procedimientos generalizados de optimización, por ejemplo, por el método de gradiente descendente.

El criterio de optimización del *error de clasificación mínimo* busca minimizar la probabilidad de error de clasificación a través de una representación suavizada, dada una función de pérdida. La medida del error de clasificación, que representa una medida de la distancia

entre la probabilidad de una decisión correcta y otras decisiones, se define como

$$d_\nu(\varphi) = -g_\nu(\varphi|\mu_C) + \log\left\{\frac{1}{V-1} \sum_{w, w \neq \nu} \exp(g_w(\varphi|\mu_C)\eta)\right\}^{1/\eta},$$

donde  $\eta$  es un número positivo y  $g_\nu = P(\varphi|\mu_\nu)$ ,  $\nu = 1, \dots, V$ . Esta medida de error es una función continua de los parámetros  $\mu_C$  y se debe asimilar con una regla de decisión. La medida  $d_\nu(\varphi)$  se usa como variable dentro de una función suavizada

$$f_\nu(\varphi|\mu_C) = f_\nu(d_\nu(\varphi)|\mu_C) = \frac{1}{1 + \exp(-\rho d_\nu(\varphi) + o_f)}$$

con  $o_f$  igual a cero y  $\rho \geq 1$ . Para cualquier sucesión desconocida  $\varphi^\nu$ , el desempeño del sistema se mide mediante  $f(\varphi^\nu|\mu_C) = \sum_{\nu=1}^V f_\nu(\varphi^\nu|\mu_C)$  que se usa como criterio de optimización en el algoritmo de gradiente descendente. Otros criterios de optimización de los parámetros se basan en la minimización de la razón del error empírico [35] y la minimización del error cuadrático medio.

La segunda forma de aproximación consiste en asumir que la señal no fue generada necesariamente por una fuente Markoviana, pero se ajusta a ciertas de sus restricciones, por ejemplo, tiene función de correlación positiva definida. El objetivo del procedimiento de diseño, consiste entonces en hallar los parámetros del modelo oculto que minimicen la información de discriminación o la entropía mutua entre el conjunto de FDP válidas  $Q$ , que satisfacen los valores medidos de señal, y el conjunto de FDP de los modelos ocultos,  $P_\mu$ . La información discriminante entre el conjunto  $Q$  y  $P_\mu$  se determina como:

$$D(Q|P_\mu) = \int q(\varphi) \ln(q(\varphi)/p(\varphi)) d\varphi \quad (4.63)$$

donde  $q$  y  $p$  son las funciones de densidad de probabilidad que corresponden a  $Q$  y  $P_\mu$ . Aunque las técnicas de optimización para (4.63) resultan ser relativamente complejas, se suele emplear el algoritmo de Baum adecuado para el caso de modelos ocultos de Markov.

**FDP de la duración de los estados.** En un modelo oculto de Markov, la FDP de duración inherente asociada con el estado  $\vartheta_i$  y coeficiente de transición propio  $\pi_{ii}$ , se asume de la forma  $p_i(d) = (\pi_{ii})^{d-1}(1 - \pi_{ii})$ , que para la mayoría de las señales físicas es inapropiada. En lugar, se prefiere modelar explícitamente la FDP de duración en forma analítica. En este caso, es necesario reformular las expresiones de actualización de los parámetros. Así, la variable de propagación directa  $\alpha_n(i)$  se define como,

$$\alpha_n(i) = P(\varphi_1, \varphi_2, \dots, \varphi_n, \vartheta_i \text{ termina en } n|\mu)$$

Se asume que se han visitado un total de  $r$  estados durante las primeras  $n$  observaciones y se denotan los estados como  $\theta_1, \theta_2, \dots, \theta_r$ , con duraciones asociadas con cada estado iguales a  $d_1, d_2, \dots, d_r$ . Las restricciones para  $\alpha_n(i)$  están dadas en la forma,  $\theta_n = \vartheta_i$  y  $\sum_{s=1}^r d_s = n$ .

La probabilidad  $\alpha_n(i)$  puede escribirse como

$$\alpha_n(i) = \sum_{\theta} \sum_d p_{\theta_1} p_{\theta_1}(d_1) P(\varphi_1, \varphi_2, \dots, \varphi_{d_1} | \theta_1) \pi_{\theta_1 \theta_2} p_{\theta_2}(d_2) P(\varphi_{d_1+1}, \dots, \varphi_{d_1+d_2} | \theta_2) \dots \\ \times \pi_{\theta_{r-1} \theta_r} p_{\theta_r}(d_r) P(\varphi_{d_1+d_2+\dots+d_{r-1}+1}, \dots, \varphi_n | \theta_r),$$

donde la suma se toma sobre todos los estados  $\theta$  y todas las posibles duraciones  $d$ . De manera recursiva,  $\alpha_n(i)$  se escribe como

$$\alpha_n(j) = \sum_{i=1}^{n_{\vartheta}} \sum_{d=1}^D \alpha_{n-d}(i) \pi_{ij} p_j(d) \prod_{s=n-d+1}^n p_{i\varphi_s}, \quad (4.64)$$

donde  $D$  es la duración máxima en un estado. Para inicializar el cálculo de  $\alpha_n(j)$  se usa

$$\alpha_1(i) = p_{\theta_1}(i) p_i(1) p_{i\varphi_1} \quad \alpha_2(i) = p_{\theta_1}(i) p_i(2) \prod_{s=1}^2 p_{i\varphi_s} + \sum_{\substack{j=1 \\ j \neq i}}^{n_{\vartheta}} \alpha_1(j) \pi_{ji} p_i(1) p_{i\varphi_2} \\ \alpha_3(i) = p_{\theta_1}(i) p_i(3) \prod_{s=1}^3 p_{i\varphi_s} + \sum_{d=1}^2 \sum_{\substack{j=1 \\ j \neq i}}^{n_{\vartheta}} \alpha_{3-d}(j) \pi_{ji} p_i(d) \prod_{s=4-d}^3 p_{i\varphi_s}$$

De igual forma, se calcula  $\alpha_4(i)$  hasta  $\alpha_D(i)$ . Finalmente, se usa (4.64) para  $n > D$ . La probabilidad  $P(\varphi|\mu)$  se calcula usando (4.44). A fin de dar expresiones de actualización en un modelo oculto de Markov de duración variable, se definen las variables:

$$\alpha_n^*(i) = P(\varphi_1, \varphi_2, \dots, \varphi_n, \vartheta_i \text{ empieza en } n+1 | \mu) \\ \beta_n(i) = P(\varphi_{n+1}, \varphi_{n+2}, \dots, \varphi_{n_{\varphi}} | \vartheta_i \text{ termina en } n, \mu) \\ \beta_n^*(i) = P(\varphi_{n+1}, \varphi_{n+2}, \dots, \varphi_{n_{\varphi}} | \vartheta_i \text{ empieza en } n+1, \mu)$$

Las relaciones entre  $\alpha, \alpha^*, \beta$  y  $\beta^*$  son

$$\alpha_n^*(i) = \sum_{j=1}^{n_{\vartheta}} \alpha_n(i) \pi_{ij}, \quad \alpha_n(i) = \sum_{d=1}^D \alpha_{n-d}^*(i) p_i(d) \prod_{s=n-d+1}^n p_{i\varphi_s} \\ \beta_n(i) = \sum_{j=1}^{n_{\vartheta}} \pi_{ij} \beta_n^*(j) \quad \beta_n^*(i) = \sum_{d=1}^D \beta_{n+d}(i) p_i(d) \prod_{s=n+1}^{n+d} p_{i\varphi_s}$$

Con base en las definiciones y relaciones anteriores, las expresiones de actualización para un modelo variable se determinan como:

$$\tilde{p}_{\theta_1}(i) = \frac{p_{\theta_1}(i) \beta_1^*(i)}{P(\varphi|\mu)}, \quad \tilde{\pi}_{ij} = \frac{\sum_{n=1}^{n_{\varphi}} \alpha_n(i) \pi_{ij} \beta_n^*(j)}{\sum_{j=1}^{n_{\vartheta}} \sum_{n=1}^{n_{\varphi}} \alpha_n(i) \pi_{ij} \beta_n^*(j)}$$



ademàs

$$\begin{aligned}
 \tilde{P}_{ik} &= \frac{\sum_{n=1}^{n_\varphi} \left( \sum_{d < n} \alpha_d^*(i) \beta_d^*(i) - \sum_{d < n} \alpha_d(i) \beta_d(i) \right) \delta(\varphi_n, v_k)}{\sum_{k=1}^{n_\nu} \sum_{n=1}^{n_\varphi} \left( \sum_{d < n} \alpha_d^*(i) \beta_d^*(i) - \sum_{d < n} \alpha_d(i) \beta_d(i) \right) \delta(\varphi_n, v_k)} \\
 \tilde{p}_i(d) &= \frac{\sum_{n=1}^{n_\varphi} \alpha_n^*(i) p_i(d) \beta_{n+d}(i) \prod_{s=n+1}^{n+d} p_{i\varphi_s}}{\sum_{d=1}^D \sum_{n=1}^{n_\varphi} \alpha_n^*(i) p_i(d) \beta_{n+d}(i) \prod_{s=n+1}^{n+d} p_{i\varphi_s}} \quad (4.68)
 \end{aligned}$$

Al introducir la FDP para la duración de los estados, para algunos problemas, la calidad del modelo mejora significativamente. Sin embargo, aparecen algunos inconvenientes tales como un incremento considerable en la carga computacional y la necesidad de estimar un número adicional de parámetros,  $D$ , por estado, lo cual requiere un mayor número de observaciones. Una posible solución consiste en usar una FDP de duración de estado paramétrica, en lugar de la no paramétrica de la ec. (4.68). En [36], se propone la  $\Gamma$ -densidad para representar la FDP de duración de los estados,  $p_i(d) = \frac{\eta_i^{\nu_i} d^{\nu_i-1} e^{-\eta_i d}}{\Gamma(\nu_i)}$ ,  $d > 0$ , con parámetros  $\nu_i$  y  $\eta_i$ , media  $\nu_i/\eta_i$  y varianza  $\nu_i/\eta_i^2$ . Las expresiones de actualización para  $\nu_i$  y  $\eta_i$ , están dadas por

$$\tilde{\eta}_j = \frac{\nu_j \sum_{n=1}^{n_\varphi} \alpha_n(j) \beta_n(j)}{\sum_{n=1}^{n_\varphi} \sum_{d \leq n} d \sum_{\substack{i=1 \\ i \neq j}}^{n_\vartheta} \alpha_{(n-d)}(i) \pi_{ij} p_j(d) \prod_{s=1}^d p_{j\varphi_{n-d+s}} \beta_n(j)} \quad (4.69)$$

$$F(\tilde{\nu}_j) = \frac{\sum_{n=1}^{n_\varphi} \alpha_n(j) \beta_n(j)}{\sum_{n=1}^{n_\varphi} \sum_{d \leq n} \log(\eta_j d) \sum_{\substack{i=1 \\ i \neq j}}^{n_\vartheta} \alpha_{(n-d)}(i) \pi_{ij} p_j(d) \prod_{s=1}^d p_{j\varphi_{n-d+s}} \beta_n(j)}, \quad (4.70)$$

donde  $F(\cdot)$  es la función digamma, la cual es diferenciable continuamente y puede obtenerse en forma de serie de potencia, entonces, (4.70) se resuelve numéricamente para  $\tilde{\nu}_j$ .

En las expresiones (4.69) y (4.70),  $\alpha_n(j)$  y  $\beta_n(j)$  están dados por

$$\begin{aligned}
 \alpha_n(j) &= \sum_{d \leq n} \sum_{\substack{i=1 \\ i \neq j}}^{n_\vartheta} \alpha_{(n-d)}(i) \pi_{ij} p_j(d) \prod_{s=1}^d p_{j\varphi_{n-d+s}}, \\
 \beta_n(j) &= \sum_{d \leq n} \sum_{\substack{j=1 \\ j \neq i}}^{n_\vartheta} \pi_{ij} p_j(d) \prod_{s=1}^d p_{j\varphi_{n+s}} \beta_{n+d}(j)
 \end{aligned}$$

**Ejercicio en el CP 4.3.** Sean dos procesos aleatorios, cada uno de los cuales corresponde al resultado de pasar RBG por un sistema lineal invariante en el tiempo con diferente función de transferencia del tipo respuesta a impulso infinita, (1.56):

$$H_0(z) = \frac{z+1}{2z}, \quad H_1(z) = \frac{1}{2}(1-z^{-1})$$

El proceso aleatorio generado, tiene un número de sucesiones  $N_s$ , y cada sucesión contiene un número de observaciones  $T$ .

```
% Generación del proceso RBG
data = randn(0*T*Ns,1);
% Definición del sistema lineal invariante en el tiempo H0
% Paso del RBG a través de H0
num0 = [1 1]; den0 = [2]; data0tem = filter(num0,den0,data);
% Definición del sistema lineal invariante en el tiempo H1
% Paso del RBG a través de H1
num1 = [1 -1]; den1 = [2 0]; data1tem = filter(num1,den1,data);
% Acondicionamiento de las sucesiones para ser utilizadas en el
% entrenamiento de los modelos ocultos de Markov
for i=1:length(data)/(0*T)
    data0T(i,:) = data0tem((i-1)*0*T+1:i*0*T)';
    data1T(i,:) = data1tem((i-1)*0*T+1:i*0*T)';
    for j = 1:size(data0T,2)/T
        data0(:,j,i) = data0T(i,(j-1)*T+1:j*T)';
        data1(:,j,i) = data1T(i,(j-1)*T+1:j*T)';
    end
end
```

A partir de los procesos generados en el paso anterior, es posible inicializar los parámetros de los modelos ocultos de Markov, de igual forma como se realizó en el ejercicio PC 4.2.

```
% Inicialización de los parámetros de los modelos ocultos de Markov para la clase 0
prior0 = normalise(rand(Q,1)); transmat0 = mk_stochastic(rand(Q,Q));
[mu0,Sigma0]=mixgauss_init(Q*M,data0, cov_type);
% Restricciones de las medidas estadísticas
mu0 = reshape(mu0, [0 Q M]); Sigma0 = reshape(Sigma0, [0 0 Q M]);
mixmat0 = mk_stochastic(rand(Q,M));
% Entrenamiento del modelo para la clase 0
[LL, priorC0, transmatC0, muC1, SigmaC0, mixmatC0] = mhmm_em(data0, prior0, ...
transmat0,mu0,Sigma0, mixmat0,'max_iter',5);
% Inicialización de los parámetros de los modelos ocultos de Markov para la clase 1
[mu0, Sigma0] = mixgauss_init(Q*M, data1, cov_type);
% Restricciones de las medidas estadísticas
mu0 = reshape(mu0, [0 Q M]); Sigma0 = reshape(Sigma0, [0 0 Q M]);
% Entrenamiento del modelo para la clase 1
[LL, priorC1, transmatC1, muC1, SigmaC1, mixmatC1] = mhmm_em(data0, prior0, ...
transmat0,mu0,Sigma0, mixmat0, 'max_iter',5);
```

Si se tiene una trayectoria nueva, se calcula la probabilidad a posteriori de que cada uno de los modelos genere la sucesión dada. La sucesión es asignada a la clase del modelo que genere la mayor probabilidad.

```
%se asume que la muestra a reconocer es de la clase 1
data3 = data1; loglik0 = mhmm_logprob(data3, priorC0, transmatC0, muC0, SigmaC0, ...
mixmatC0);
loglik1 = mhmm_logprob(data3, priorC1, transmatC1, muC1, SigmaC1, ...
mixmatC1);
if loglik1 > loglik0
    class = 1;
else
    class = 0;
end
```

**Comparación de los modelos de ocultos de Markov.** Dados dos procesos aleatorios de Markov de primer orden  $\mu_1$  y  $\mu_2$  se debe hallar una medida de similitud o distancia entre los mismos,  $d(\mu_1, \mu_2)$ , con el fin de medir su equivalencia estadística. En la práctica, se han propuesto diversas medidas de distancia. La primer medida propuesta corresponde a una forma generalizada de la distancia euclídea entre las matrices de probabilidad estado-observación, que se define como:

$$d(\mu_1, \mu_2) = \sqrt{\frac{1}{n_\vartheta} \sum_{i=1}^{n_\vartheta} \sum_{k=1}^{n_v} \|P_{ik}^{(1)} - P_{ik}^{(2)}\|^2}, \quad (4.71)$$

Un caso más general de esta medida se realiza encontrando el estado del segundo proceso que minimiza la diferencia entre las probabilidades del los modelos,

$$d(\mu_1, \mu_2) = \left\{ \frac{1}{n_\vartheta n_v} \sum_{i=1}^{n_\vartheta} \sum_{k=1}^{n_v} \left( P_{ik}^{(1)} - P_{c(i)k}^{(2)} \right)^2 \right\}^{1/2} \quad (4.72)$$

donde  $c(i)$  es la permutación de los estados que minimiza (4.72). Las medidas (4.71) y (4.72) son inadecuadas, dado que no toman en cuenta la estructura temporal representada en la cadena de Markov, por lo que podría darse el caso en el cual es posible encontrar un par de modelos ocultos de Markov, con una distancia entre sí que tienda a cero, pero con medidas respectivas de probabilidad,  $P_\mu$  y  $P_{\mu'}$ , completamente diferentes.

Una medida alterna, que define la distancia de *Kullback - Leibler* entre las probabilidades (4.44),  $P(\cdot|\mu_i)$ , está dada por

$$d(\mu_1, \mu_2) = \frac{1}{2} (\log(P_{11}P_{22}) - \log(P_{12}P_{21})), \quad (4.73)$$

donde

$$P_{ij} = P(\varphi_i|\mu_j)^{1/n_{\varphi_i}}, \quad (4.74)$$

siendo  $n_{\varphi_i}$  la longitud de la sucesión  $\varphi_i$  generada de forma estocástica a partir del modelo  $\mu_i$ . La medida  $d(\mu_1, \mu_2)$  está determinada unívocamente sólo en el límite, cuando  $n_{\varphi_i} \rightarrow \infty$ . Así mismo, se puede demostrar que, si  $P_{ii}$  es un máximo global, la distancia de *Kullback-Leibler* tiene las siguientes propiedades [37]:

1.  $d(\mu_1, \mu_2) = d(\mu_2, \mu_1)$
2.  $d(\mu_1, \mu_2) \geq 0$
3.  $d(\mu_1, \mu_2) = 0$  si  $\mu_1 \sim \mu_2$  ó  $\varphi_1 = \varphi_2$

Una de las desventajas que tiene este tipo de distancia se encuentra en la necesidad de realizar la simulación, por ejemplo por Monte Carlo, para generar las sucesiones  $\hat{\varphi}_i$ , lo cual eleva el costo computacional. En [38], se propone una cota superior para la distancia de Kullback Leibler, que hace innecesario el procedimiento de simulación.

Otras distancias entre modelos ocultos de Markov se definen sobre el concepto de *probabilidad de co-emisión* [39],  $\sum_{\varphi} \mu_1(\varphi) \mu_2(\varphi)$ , que es la probabilidad de que independientemente los modelos generen la misma sucesión. Esta probabilidad requiere que el modelo del proceso de Markov tenga la estructura de un modelo de nacimiento. El cálculo de la probabilidad de co-emisión implica la construcción de una tabla  $\mathbf{A}$  indexada por los estados de ambos modelos ocultos de Markov, tal que la entrada  $P_{\mathbf{A}}(\theta_{\mu_1}, \theta_{\mu_2})$  de la tabla (donde  $\theta_{\mu_i}$  es un estado de  $\mu_i$ ,  $i = 1, 2$ ) representa la probabilidad de encontrarse en el estado  $\theta_{\mu_1}$  en  $\mu_1$  y  $\theta_{\mu_2}$  en  $\mu_2$ , además de haber generado independientemente sucesiones idénticas de observación en las rutas de estados, que conducen a  $\theta_{\mu_1}$  y a  $\theta_{\mu_2}$ . El valor de la tabla correspondiente a los dos estados finales de cada modelo es la probabilidad de co-emisión.

Basados en la probabilidad de co-emisión de dos modelos dados,  $P_{\mathbf{A}}(\mu_1, \mu_2)$ , se definen las siguientes dos métricas:

$$d_{ang}(\mu_1, \mu_2) = \arccos(P_{\mathbf{A}}(\mu_1, \mu_2) / \sqrt{P_{\mathbf{A}}(\mu_1, \mu_1) P_{\mathbf{A}}(\mu_2, \mu_2)})$$

$$d_{dif}(\mu_1, \mu_2) = \sqrt{P_{\mathbf{A}}(\mu_1, \mu_1) + P_{\mathbf{A}}(\mu_2, \mu_2) - 2P_{\mathbf{A}}(\mu_1, \mu_2)}$$

Otra versión de métrica se conoce como la *medida de similitud* que mide la semejanza entre dos trayectorias estocásticas con dimensión múltiple en correspondencia con los modelos ocultos de Markov dados,

$$d(\varphi_1, \varphi_2) = \sqrt{\frac{P_{21}P_{12}}{P_{11}P_{22}}}$$

donde,  $P_{ij} = P(\varphi_i | \mu_j)^{n_{\varphi_i}}$ , representa la probabilidad de la sucesión de observación  $\varphi_i$  dada por el modelo  $\mu_j$ , normalizado con respecto a  $n_{\varphi_i}$ , donde  $n_{\varphi_i}$  es la longitud de la sucesión  $\varphi_i$ . Se puede demostrar [40] que si  $P_{ii}$  es un máximo global, la medida de similitud tiene las siguientes propiedades:

1.  $d(\varphi_1, \varphi_2) = d(\varphi_2, \varphi_1)$
2.  $0 < d(\varphi_1, \varphi_2) \leq 1$
3.  $d(\varphi_1, \varphi_2) = 1$  si  $\mu_1 \sim \mu_2$  ó  $\varphi_1 = \varphi_2$

En algunas ocasiones es más conveniente representar la similitud entre dos modelos de Markov a través de una medida de distancia en lugar de una medida de similitud. Dada la medida de similitud  $d(\varphi_1, \varphi_2)$ , la medida de distancia se puede obtener a partir de

$$d(\varphi_1, \varphi_2) = -\log d(\varphi_1, \varphi_2) \quad (4.75)$$

tal que se cumple  $d(\varphi_1, \varphi_2) = d(\varphi_2, \varphi_1)$ ,  $d(\varphi_1, \varphi_2) \geq 0$  y  $d(\varphi_1, \varphi_2) = 1$  si  $\mu_1 \sim \mu_2$  ó  $\varphi_1 = \varphi_2$ .

A diferencia de las sucesiones de observación  $\varphi_i$ , las sucesiones de observación  $\widehat{\varphi}_i$  de (4.74) no son únicas debido a que son generadas de forma estocástica a partir de  $\widehat{\mu}_i$ .

Aunque de forma general  $d(\hat{\mu}_1, \hat{\mu}_2)$  (ecuación (4.73)) y  $d(\varphi_1, \varphi_2)$  (ecuación (4.75)) no son equivalentes, bajo ciertas presunciones las dos nociones (distancia entre los modelos ocultos y distancia entre sucesiones de observación) convergen a una equivalencia. Específicamente, se tiene que,  $d(\varphi_1, \varphi_2) = d(\hat{\mu}_1, \hat{\mu}_2)$  si y sólo si

1.  $\mu_1 \sim \hat{\mu}_1$  y  $\mu_2 \sim \hat{\mu}_2$
2.  $P_{11}$  y  $P_{22}$  son máximos globales.
3.  $\hat{n}_{\varphi_i} \rightarrow \infty$

**Ejercicio en el CP 4.4.** Dados dos modelos ocultos de Markov, calcular la distancia de Kullback-Leibler entre los modelos. Para realizar este ejercicio, serán tomados los modelos que fueron entrenados a partir de procesos aleatorios diferentes, en el ejercicio PC 4.3. El modelo generado a partir de las sucesiones de observaciones 0, es llamado modelo clase  $C_0$ . De igual forma, el modelo generado a partir de las sucesiones de observaciones 1, es llamado modelo clase  $C_1$ . Esta notación se mantiene para cada uno de los parámetros del modelo.

En primer lugar es necesario definir la longitud de las sucesiones de observación. Debe tenerse en cuenta, que la medida de distancia Kullback-leibler está definida para una longitud de la sucesión de observaciones tendiente a infinito, por lo tanto, es necesario determinar una longitud lo suficientemente grande. En este caso, se asume una longitud  $T = 10000$ .

```
% Implementación de la distancia de Kullback - Leibler
% Definición de la longitud de las sucesiones de observación.
T = 10000;      %la longitud de las secuencias debe tender a inf
nseq = 1;      % número de secuencias
```

Después de definir la longitud de las sucesiones y el número de sucesiones, es necesario generar las sucesiones de observación a partir de los modelos dados y subsecuentemente se calculan las probabilidades definidas en la ecuación (4.74).

```
% Generación de las sucesiones a partir de los modelos dados
% Generación de sucesiones a partir del modelo C0
[obs0, hidden0] = mhmm_sample(T, nseq, priorC0, transmatC0, muC0, SigmaC0, mixmatC0);
% Generación de sucesiones a partir del modelo C1
[obs1, hidden1] = mhmm_sample(T, nseq, priorC1, transmatC1, muC1, SigmaC1, mixmatC1);
%cálculo de las probabilidades Pij
loglik00 = mhmm_logprob(obs0, priorC0, transmatC0, muC0, SigmaC0, mixmatC0);
loglik01 = mhmm_logprob(obs0, priorC1, transmatC1, muC1, SigmaC1, mixmatC1);
loglik10 = mhmm_logprob(obs1, priorC0, transmatC0, muC0, SigmaC0, mixmatC0);
loglik11 = mhmm_logprob(obs1, priorC1, transmatC1, muC1, SigmaC1, mixmatC1);
```

Según la definición (4.74), las medidas de probabilidad  $P_{ij}$  están normalizadas con respecto a la longitud de las sucesiones de observación, por lo tanto, antes de calcular la medida de distancia se debe realizar la normalización y posteriormente aplicar la definición (4.73), para así calcular la medida de distancia.

```
%Normalización
loglik00 = loglik00/T; % Se hace lo mismo para las demás f. de verosimilitud
% Cálculo de la medida de distancia
d12=loglik11-loglik12; d21=loglik22-loglik21;
% Cálculo de la distancia simétrica
Ds=(d12+d21)/2;
```

## Problemas

**Ejercicio en el CP 4.5.** Realice un análisis comparativo de diferentes modelos ocultos de Markov entrenados con el mismo conjunto de datos, variando el número de estados y de mezclas Gaussianas en el modelo.

**Ejercicio en el CP 4.6.** Calcular la secuencia de estados más probable para una sucesión y un modelo dado, empleando las funciones:

```
obslik = mixgauss_prob(Secuencia, mu, Sigma, mixmat);  
[path,loglik] = viterbi_path(prior, transmat,obslik, 0);
```

El parámetro loglik, en este caso, contiene la probabilidad conjunta de la sucesión de observación y la sucesión de estados. Realizar el análisis comparativo de ésta probabilidad conjunta y la probabilidad máximo a posteriori, para diferentes modelos variando el número de estados y de mezclas en el modelo.

**Ejercicio en el CP 4.7.** Tomar como base la implementación de la distancia de Kullback- Leibler desarrollada en ejemplo 4.4, e implementar la medida de distancia por similitud (ecuación (4.75)), compare el comportamiento de ambas medidas de distancia para un número diferente de estados y de mezclas en el modelo.

# Capítulo 5

## Transformación de procesos aleatorios

En general, el paso de procesos aleatorios por circuitos lineales consiste en el análisis de las propiedades de aleatoriedad de los procesos de entrada y salida del sistema, dada su forma de transformación. Por regla, la solución general es difícil de alcanzar, por lo tanto, ésta se simplifica hasta hallar las funciones de los momentos para el caso no estacionario, o de los valores de los momentos y la función de correlación (o su respectiva densidad espectral de potencia) para el caso estacionario.

### 5.1. Paso de señales aleatorias por sistemas lineales

En concordancia con la descripción de sistemas lineales hecha en el numeral §1.2.3, se presentan los siguientes tres métodos de análisis.

#### 5.1.1. Análisis en el tiempo

La forma directa de análisis es la convolución (1.39), que relaciona la señal aleatoria de salida  $\eta(t)$  con la entrada  $\xi(t)$ , a través de la respuesta a impulso del sistema  $h(t)$ :

$$\eta(t) = \int_0^{\infty} \xi(t - \tau) h(\tau) d\tau \quad (5.1)$$

El límite superior de la integral en (5.1) es escogido igual al valor  $\tau \rightarrow \infty$ , a fin de analizar el comportamiento asintótico del sistema en régimen de estado estable, antes que hacerlo igual a  $t$ , que corresponde al régimen transitorio.

El operador de promedio sobre ambos términos de (5.1) se generaliza en la forma:

$$E\left\{\int_T \xi(t) g(t) dt\right\} = \int_T E\{\xi(t)\} g(t) dt$$

donde  $g(t)$  es una función no aleatoria que varía en el tiempo.

El intercambio en la operación de integración con el operador de promedio, implica la simplificación del análisis; intercambio que se realiza cuando se cumplen las condiciones:

$$(a). \int_T E \{ |\xi(t)| \} g(t) dt < \infty$$

Al asumir la estacionariedad de los procesos de entrada, cuando los operadores de promedio son invariantes en el tiempo (ver (3.1)), entonces, se consideran una constante  $k$  que sale fuera de la integral. En calidad de función  $g(t)$  representativa del sistema se emplea la respuesta a impulso  $h(t)$ , por lo que la restricción, que corresponde a la condición de estabilidad del sistema, se resume a  $k \int_T |h(t)| dt < \infty$ .

$$(b). \xi(t) \in T, \text{ aunque se puede tener que } T \rightarrow \infty.$$

**Valor medio de salida.** Para un sistema lineal e invariante en el tiempo, cuando a su entrada se tiene una señal estacionaria, teniendo en cuenta (5.1), está dado por

$$\begin{aligned} E \{ \eta(t) \} &= E \left\{ \int_0^\infty \xi(t-\lambda) h(\lambda) d\lambda \right\} = \int_0^\infty E \{ \xi(t-\lambda) \} h(\lambda) d\lambda \\ &= m_{1\xi} \int_0^\infty h(\lambda) d\lambda \end{aligned} \quad (5.2)$$

Sin embargo, la integral de la respuesta a impulso en (5.2) se puede representar de forma alterna como la TF para el caso único de análisis de la componente  $\omega = 0$ , esto es,

$$\begin{aligned} E \{ \eta(t) \} &= m_{1\xi} \int_0^\infty h(\lambda) d\lambda = m_{1\xi} \int_0^\infty h(\lambda) e^{-j\omega\lambda} d\lambda \Big|_{\omega=0} \\ &= m_{1\xi} H(0) \end{aligned} \quad (5.3)$$

La expresión (5.3) es evidente, porque muestra que el primer momento, al ser una constante, afecta solamente la componente  $\omega = 0$  de la función de transferencia del sistema. Sin embargo, un proceso aleatorio a la salida de un circuito lineal, activado por una entrada a partir de  $t = 0$ , se expresa por la integral (1.39), que se diferencia de la (5.1) por su límite superior. Por lo tanto, la expresión (5.2) para el valor medio se corrige en la forma:

$$\begin{aligned} E \{ \eta(t) \} &= m_{1\xi} \int_{-\infty}^t h(\lambda) d\lambda = m_{1\xi} a(t) \\ &= m_{1\eta}(t) \end{aligned} \quad (5.4)$$

El valor medio para cualquier sistema lineal resulta directamente proporcional a la función de transición del sistema  $a(t)$ , que a su vez depende del tiempo, entonces, el proceso a la salida es no estacionario, mientras el sistema no esté en régimen de estado estable.



**Ejemplo 5.1.** En el circuito  $RC$  del ejemplo 1.3, se tiene una señal aleatoria  $\xi(t)$  de voltaje, definida a partir del momento  $t = 0$  que corresponde a RBG con valor medio  $m_{1\xi}$ . Determinar el cambio del valor medio en el tiempo del voltaje de salida  $\eta(t)$

La ecuación diferencial del circuito  $RC$  de la Figura 1.4(a) tiene la forma:

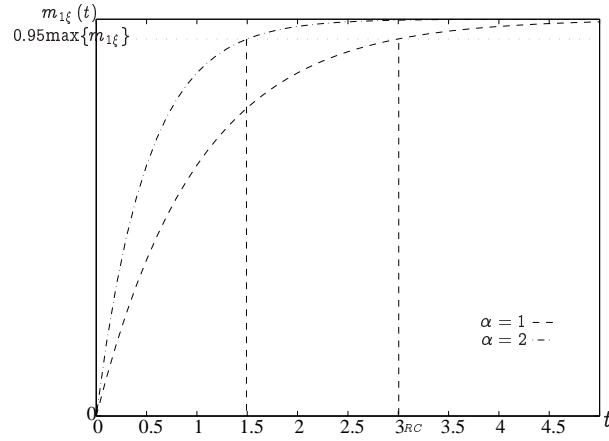
$$\frac{d\eta}{dt} + \alpha\eta(t) = \alpha\xi(t), \quad \alpha = 1/RC$$

Considerando que el capacitor en el momento  $t = 0$  está descargado, y por ende,  $\eta(0) = 0$ , la anterior ecuación diferencial tiene la siguiente solución:

$$\eta(t) = \alpha e^{-\alpha t} \int_0^t e^{\alpha\lambda} \xi(\lambda) d\lambda$$

Por lo que el valor medio del voltaje de salida  $\eta(t)$  del circuito  $RC$  es igual a:

$$E\{\eta(t)\} = \alpha e^{-\alpha t} \int_0^t e^{\alpha\lambda} E\{\xi(\lambda)\} d\lambda = \alpha e^{-\alpha t} \int_0^t e^{\alpha\lambda} m_{1\xi} d\lambda = m_{1\xi} \alpha e^{-\alpha t} (e^{\alpha t} - 1)$$



**Figura 5.1.** Cambio del valor medio en el tiempo

De esta manera, el valor medio a la salida cambia en el tiempo de la siguiente forma:

$$m_{1\eta}(t) = m_{1\xi} \alpha e^{-\alpha t} (e^{\alpha t} - 1)$$

La Figura 5.1 representa el cambio del valor medio (5.4) de la señal aleatoria de salida  $\eta(t)$  en el tiempo (para dos valores diferentes de  $\alpha$ ), que corresponde a un intervalo de no estacionariedad del proceso. Este intervalo es transitorio y su existencia, siendo proporcional al factor de inercia  $\alpha = RC$  del sistema, se considera insignificante con relación al intervalo total de análisis del proceso.

Cuando a la entrada del sistema lineal se tiene un proceso no estacionario, determinado a partir del momento  $t = 0$ , el valor medio del proceso a la salida, dado por la relación (5.2), es igual a:

$$E \{ \eta(t) \} = \int_0^t h(\lambda) m_{1\xi}(t - \lambda) d\lambda = h(t) * m_{1\xi}(t) \quad (5.5)$$

La expresión (5.5) coincide con el resultado obtenido en (3.6a), para el caso de transformación lineal de señales aleatorias.

En general, el análisis en estado transitorio del sistema lineal se puede realizar para cada uno de los momentos, tanto en el caso estacionario, como en el no estacionario. Sin embargo, debido a la dificultad en la solución del caso transitorio, es frecuente que el análisis se limite al régimen de estado estable.

**Valor cuadrático medio de salida.** El análisis es similar al caso anterior del valor medio de la señal aleatoria de salida, aunque, aquí es preferible introducir dos variables de integración, a fin de representar la multiplicación interna del operador de promedio en forma de una integral doble, de la siguiente forma:

$$\begin{aligned} E \{ \eta^2(t) \} &= E \left\{ \int_0^\infty \xi(t - \lambda_1) h(\lambda_1) d\lambda_1 \int_0^\infty \xi(t - \lambda_2) h(\lambda_2) d\lambda_2 \right\} \\ &= E \left\{ \int_0^\infty d\lambda_1 \int_0^\infty \xi(t - \lambda_1) \xi(t - \lambda_2) h(\lambda_1) h(\lambda_2) d\lambda_2 \right\} \\ &= \int_0^\infty d\lambda_1 \int_0^\infty E \{ \xi(t - \lambda_1) \xi(t - \lambda_2) \} h(\lambda_1) h(\lambda_2) d\lambda_2 \end{aligned} \quad (5.6)$$

La operación de promedio dentro de la integral corresponde a la función de correlación de la señal aleatoria de entrada, evaluada en los siguientes puntos:

$$\begin{aligned} E \{ \xi(t - \lambda_1) \xi(t - \lambda_2) \} &= R_\xi(t - \lambda_1 - t + \lambda_2) \\ &= R_\xi(\lambda_2 - \lambda_1) \end{aligned}$$

Por lo que la expresión (5.6) se formula en la siguiente forma:

$$E \{ \eta^2(t) \} = \int_0^\infty d\lambda_1 \int_0^\infty R_\xi(\lambda_2 - \lambda_1) h(\lambda_1) h(\lambda_2) d\lambda_2 \quad (5.7)$$

---

**Ejemplo 5.2.** En el circuito  $RC$  del ejemplo 1.3, se tiene a la entrada una señal aleatoria  $\xi(t)$  de voltaje, que corresponde a RBG. Determinar el valor cuadrático medio del voltaje de salida  $\eta(t)$ .

El valor cuadrático medio de salida determinado por (5.7), teniendo en cuenta el valor de la función de correlación del RBG (3.34), es igual a,

$$\begin{aligned} E\{\eta^2(t)\} &= \int_0^\infty d\lambda_1 \int_0^\infty \frac{N_0}{2} \delta(\lambda_2 - \lambda_1) h(\lambda_1) h(\lambda_2) d\lambda_2 = \frac{N_0}{2} \int_0^\infty h^2(\lambda) d\lambda \\ &= \frac{N_0}{2} \int_0^\infty (\alpha e^{-\alpha\lambda})^2 d\lambda = \frac{N_0 \alpha^2}{2} \int_0^\infty e^{-2\alpha\lambda} d\lambda = \frac{N_0 \alpha^2}{2} \left. \frac{e^{-2\alpha\lambda}}{-2\alpha} \right|_0^\infty \\ &= \frac{N_0}{2} \frac{\alpha}{2} \end{aligned}$$

En consideración, cuando en vez de asumir un ruido con varianza infinita, se asume un ruido con correlación nula, valor medio cero, pero varianza finita,  $\sigma_\xi^2 = D < \infty$ , el proceso a la salida del dispositivo lineal es unívocamente cero,

$$\eta(t) = \int_0^t h(\tau) \xi(t - \tau) d\tau \equiv 0$$

en la medida en que los momentos (5.5) y (5.7) son iguales a 0.

Cabe anotar, que la resolución de este ejemplo es relativamente simple, pero solo si el proceso a la entrada es del RBG, en caso contrario, cuando el proceso es otra naturaleza, la resolución de la integral doble (5.7) presenta serias dificultades.

**Función de correlación de salida.** En este caso, se puede tomar el resultado obtenido en el análisis de la transformación de señales aleatorias con operadores lineales. Particularmente, asumiendo la estacionariedad del proceso de entrada de (3.6b), se obtiene:

$$R_\eta(\tau) = \int_0^\infty d\lambda_1 \int_0^\infty R_\xi(\lambda_2 - \lambda_1 - \tau) h(\lambda_1) h(\lambda_2) d\lambda_2 \quad (5.8)$$

La varianza del proceso a la salida resulta de hacer  $\tau = 0$  en la expresión (5.8), inclusive para el caso de análisis de regimen transitorio

$$\sigma_\eta^2(t) = \int_0^t d\lambda_1 \int_0^t R_\xi(\lambda_2 - \lambda_1) h(\lambda_1) h(\lambda_2) d\lambda_2$$

**Ejemplo 5.3.** En el circuito  $RC$  del ejemplo 1.3, se tiene una señal aleatoria  $\xi(t)$  de voltaje, definida a partir del momento  $t = 0$  que corresponde a RBG. Determinar el cambio de la función de correlación en el tiempo del voltaje de salida  $\eta(t)$ .

La expresión (5.8), sustituyendo la función de correlación del RBG (3.34), toma la forma:

$$\begin{aligned} R_\eta(\tau) &= \int_0^\infty d\lambda_1 \int_0^\infty \frac{N_0}{2} \delta(\lambda_2 - \lambda_1 - \tau) h(\lambda_1) h(\lambda_2) d\lambda_2 \\ &= \frac{N_0}{2} \int_0^\infty h(\lambda) h(\lambda + \tau) d\lambda = \frac{N_0}{2} R_h(\tau) \end{aligned}$$

De la anterior expresión, se observa que en presencia del RBG a la entrada de un sistema lineal e invariante en el tiempo, la función de correlación de la señal aleatoria de salida es proporcional a una dependencia, que de manera condicional, se puede denominar función de correlación de la respuesta a impulso del sistema, esto es,

$$R_\eta(\tau) = S_\xi(\omega) R_h(\tau)$$

El análisis en el tiempo de la función de correlación de salida se puede realizar integrando hasta el valor  $t$ ,

$$R_\eta(t_1, t_2) = \frac{N_0}{2} \alpha^2 e^{-\alpha(t_1+t_2)} \int_0^{t_1} \int_0^{t_2} e^{\alpha(\lambda_1+\lambda_2)} \delta(\lambda_1 - \lambda_2) d\lambda_2 d\lambda_1$$

Por cuanto, en el desarrollo de la integral doble es necesario tener en cuenta que es cierta la siguiente expresión:

$$\int_{\lambda_0-\varepsilon}^{\lambda_0+\varepsilon} g(\lambda_0 - \lambda) d\lambda = g(\lambda_0)$$

solo para  $\varepsilon > 0$ , cuando el punto específico  $\lambda = \lambda_0$  se encuentra dentro de los límites de integración, entonces, el análisis se lleva a cabo para dos casos:

1.  $t_2 - t_1 > 0$ :

$$\begin{aligned} \int_0^{t_1} \int_0^{t_1} e^{\alpha(\lambda_1+\lambda_2)} \delta(\lambda_1 - \lambda_2) d\lambda_2 d\lambda_1 &= \int_0^{t_1} e^{\alpha\lambda_2} d\lambda_2 \int_0^{t_1} e^{\alpha\lambda_1} \delta(\lambda_1 - \lambda_2) d\lambda_1 \\ &= \int_0^{t_1} e^{2\alpha\lambda_2} d\lambda_2 = \frac{1}{2\alpha} (e^{2\alpha t_1} - 1) \end{aligned}$$

2.  $t_2 - t_1 < 0$ :

$$\int_0^{t_2} \int_0^{t_2} e^{\alpha(\lambda_1+\lambda_2)} \delta(\lambda_1 - \lambda_2) d\lambda_2 d\lambda_1 = \frac{1}{2\alpha} (e^{2\alpha t_2} - 1)$$

Juntando ambas soluciones se obtiene la siguiente función de correlación del proceso de salida:

$$R_\eta(t, \tau) = \frac{\alpha N_0}{4} e^{\alpha|\tau|} (1 - e^{-2\alpha t})$$

La varianza resulta al hacer  $\tau = 0$ :

$$\sigma_{\eta}^2(t) = \frac{\alpha N_0}{4} (1 - e^{-2\alpha t})$$

Como se observa, la varianza tiene la mismo forma de cambio que la media a la salida del ejemplo 5.1; sólo que la varianza crece dos veces más rápido que la media, esto es, tiene un menor intervalo de transición (curva ' $\cdot$ ' para  $\alpha = 2$  de la Figura 5.1).

**Funciones de correlación mutua entrada-salida.** En los sistemas lineales, es de suponer alguna relación de dependencia estadística entre los señales aleatorias de entrada y salida, dada por la respectiva función de correlación mutua, que se considera en dos formas:

$$\begin{aligned} R_{\xi\eta}(\tau) &= E\{\xi(t)\eta(t+\tau)\} = E\left\{\xi(t)\int_0^{\infty}\xi(t+\tau-\lambda)h(\lambda)d\lambda\right\} \\ &= \int_0^{\infty} E\{\xi(t)\xi(t+\tau-\lambda)\}h(\lambda)d\lambda = \int_0^{\infty} R_{\xi}(\tau-\lambda)h(\lambda)d\lambda \end{aligned} \quad (5.9a)$$

$$\begin{aligned} R_{\eta\xi}(\tau) &= E\{\xi(t+\tau)\eta(t)\} = E\left\{\xi(t+\tau)\int_0^{\infty}\xi(t-\lambda)h(\lambda)d\lambda\right\} \\ &= \int_0^{\infty} E\{\xi(t+\tau)\xi(t-\lambda)\}h(\lambda)d\lambda = \int_0^{\infty} R_{\xi}(\tau+\lambda)h(\lambda)d\lambda \end{aligned} \quad (5.9b)$$

Por cuanto la función de convolución dentro de ambas integrales es par con relación a  $\lambda = -\tau$ , y la respuesta a impulso se define solo para valores positivos de  $\lambda$ , entonces, es de esperar que la función de correlación mutua  $R_{\xi\eta}(\tau)$ , en general, se diferencie de  $R_{\eta\xi}(\tau)$ .

**Ejemplo 5.4.** Hallar ambas funciones de correlación mutua entre la entrada y la salida del circuito  $RC$  del ejemplo 1.3, si se tiene una señal de entrada aleatoria  $\xi(t)$  de voltaje, que corresponde a RBG.

Teniendo en cuenta las expresiones (5.9a) y (5.9b), se obtiene

$$\begin{aligned} R_{\xi\eta}(\tau) &= \frac{N_0}{2} \int_0^{\infty} \delta(\tau-\lambda)h(\lambda)d\lambda = \frac{N_0}{2} h(\tau)u(\tau) \\ R_{\eta\xi}(\tau) &= \frac{N_0}{2} \int_0^{\infty} \delta(\tau+\lambda)h(\lambda)d\lambda = \frac{N_0}{2} h(-\tau)u(-\tau) \end{aligned}$$

Los resultados obtenidos sobre la asimetría de las respectivas funciones de correlación conjunta confirman lo obtenido en el numeral §3.1.4.

### 5.1.2. Análisis en la frecuencia

El análisis matemático en el tiempo presenta dificultades de solución en el desarrollo de los momentos de salida, para el caso de procesos aleatorios con FDP diferente a la Gaussiana, debido a la presencia en la mayoría de ellos de las integrales dobles, de las cuales una de ellos es la correlación. Así por ejemplo, la expresión (5.8) entre las funciones de correlación (3.6b) de entrada y salida en un sistema lineal e invariante en el tiempo, en caso de análisis del régimen estacionario y para una señal aleatoria estacionaria, está dada por la relación:

$$R_{\eta}(\tau) = \int_0^{\infty} d\lambda_1 \int_0^{\infty} R_{\xi}(\lambda_2 - \lambda_1 - \tau) h(\lambda_1) h(\lambda_2) d\lambda_2$$

En este sentido, es preferible realizar el análisis de la transformación de cada uno de los momentos y valores de aleatoriedad sobre el dominio de la frecuencia, a fin de reducir la complejidad del desarrollo matemático.

La representación en la frecuencia de los procesos aleatorios está dada por la DEP, la cual a su vez, se relaciona con la función de correlación a través de las transformadas de Wiener-Jinchin (ecs. (3.27) y (3.28)):

$$S_{\eta}(\omega) = \mathcal{F}\{R_{\eta}(\tau)\} = \int_{-\infty}^{\infty} \left\{ \int_0^{\infty} d\lambda_1 \int_0^{\infty} R_{\xi}(\lambda_2 - \lambda_1 - \tau) h(\lambda_1) h(\lambda_2) \right\} e^{-j\omega\tau} d\tau$$

Cambiando el orden de integración, se obtiene:

$$\begin{aligned} S_{\eta}(\omega) &= \int_0^{\infty} d\lambda_1 \int_0^{\infty} h(\lambda_1) h(\lambda_2) d\lambda_2 \int_{-\infty}^{\infty} R_{\xi}(\lambda_2 - \lambda_1 - \tau) e^{-j\omega\tau} d\tau \\ &= \int_0^{\infty} d\lambda_1 \int_0^{\infty} h(\lambda_1) h(\lambda_2) d\lambda_2 S_{\xi}(\omega) e^{-j\omega(\lambda_2 - \lambda_1)} d\lambda_2 \\ &= S_{\xi}(\omega) \int_0^{\infty} h(\lambda_1) e^{j\omega\lambda_1} d\lambda_1 \int_0^{\infty} h(\lambda_2) d\lambda_2 e^{-j\omega\lambda_2} d\lambda_2 \\ S_{\eta}(\omega) &= S_{\xi}(\omega) H(-\omega)H(\omega) = S_{\xi}(\omega) |H(\omega)|^2 \end{aligned} \quad (5.10)$$

La expresión (5.10) es similar a la obtenida para la DEP de señales en (1.40); analogía que puede ser empleada en los métodos de análisis de funciones de transferencia descritas por funciones racionales, cuando a la entrada se tienen procesos estacionarios.

De la expresión (3.25), se puede encontrar la varianza:

$$\sigma_{\eta}^2 = \frac{1}{2\pi} \int_{-\infty}^{\infty} S_{\eta}(\omega) d\omega = \int_{-\infty}^{\infty} S_{\xi}(\omega) H(\omega) H(-\omega) d\omega \quad (5.11)$$

El valor cuadrático medio, teniendo en cuenta la relación dada en (5.3) para el valor

medio, se halla de la siguiente manera

$$m_{2\eta} = \sigma_{\eta}^2 + m_{1\eta}^2 = \int_{-\infty}^{\infty} S_{\xi}(\omega) H(\omega) H(-\omega) d\omega + m_{1\xi}^2 (H(0))^2$$

**Ejemplo 5.5.** Hallar la DEP de salida para el circuito *RC* del ejemplo 1.3, si a la entrada se tiene RBG  $\xi(t)$ .

Teniendo en cuenta la densidad espectral de potencia del ruido blanco (3.33), entonces

$$S_{\eta}(\omega) = \frac{N_0}{2} |H(\omega)|^2 = \frac{N_0}{2} \frac{\alpha}{j\omega + \alpha} \frac{\alpha}{-j\omega + \alpha} = \frac{N_0}{2} \left| \frac{\alpha^2}{\omega^2 + \alpha^2} \right|$$

La varianza, en este caso, se halla integrando la anterior expresión:  $\sigma_{\eta}^2 = \frac{N_0}{2} \frac{\alpha}{2}$ .

**Ejemplo 5.6.** Dado un proceso estacionario  $\xi(t)$ , con valor medio  $m_{1\xi}$  y función de correlación  $R_{\xi} = \sigma_{\xi}^2 r_{\xi}(\tau)$ , a la entrada de un dispositivo de promedio deslizante con apertura finita de tiempo  $(t-T, t+T)$ . Hallar el cambio de los respectivos valores medios.

El proceso a la salida se determina como

$$\eta(t) = (2T)^{-1} \int_{t-T}^{t+T} \xi(\lambda) d\lambda \quad (1)$$

con valor medio

$$m_{1\eta} = E\{\eta(t)\} = \frac{1}{2T} \int_{t-T}^{t+T} E\{\xi(\lambda)\} d\lambda = m_{\xi}$$

y función de correlación

$$R_{\eta}(\tau) = E\{\eta(t) - m_{1\eta}, \eta(t+\tau) - m_{1\eta}\} = \frac{1}{4T^2} \int_{t-T}^{t+T} \int_{t+\tau-T}^{t+\tau+T} R_{\xi}(u-v) du dv,$$

El núcleo de la integral se reemplaza por la expresión:

$$R_{\xi}(u-v) = \frac{1}{2\pi} \int_{-\infty}^{\infty} S_{\xi}(\omega) \exp(j\omega(u-v)) d\omega$$

se obtiene

$$\begin{aligned} R_{\eta}(\tau) &= \frac{1}{2\pi 4T^2} \int_{-\infty}^{\infty} S_{\xi}(\omega) d\omega \int_{t-T}^{t+T} \exp(-j\omega v) dv \int_{t+\tau-T}^{t+\tau+T} \exp(j\omega u) du \\ &= \frac{1}{2\pi} \int_{-\infty}^{\infty} S_{\xi}(\omega) (\text{sinc}(\omega T))^2 d\omega \end{aligned} \quad (2)$$

Por lo tanto, el espectro de salida es de la forma,  $S_\eta(\omega) = S_\xi(\omega) |\text{sinc}(T\omega)|^2$ , que se concentra en el lóbulo principal de la función sinc, en la cercanía de los valores  $\omega T$ . De esta manera, se eliminan las componentes espectrales superiores, esto es, se eliminan los cambios rápidos del proceso  $\xi(t)$  en el tiempo.

La comparación de (5.1) con (1) muestra que la operación de promedio deslizando en el tiempo se analiza como el paso de una señal aleatoria  $\xi(t)$  a través de un sistema lineal con respuesta a impulso y función de frecuencia, dadas respectivamente como,

$$h(t) = \begin{cases} 1/2T, & |t| < T \\ 0, & |t| > T \end{cases}, \quad H(\omega) = \text{sinc}(\Delta\omega)$$

Reemplazando en (2),  $S_\xi(\omega) = \mathcal{F}\{R_\xi(\lambda)\}$  e intercambiando el orden de integración, se obtiene la función de correlación de salida,

$$R_\xi(\tau) = \frac{1}{2T} \int_{-2T}^{2T} \left(1 - \frac{|\lambda|}{2T}\right) R_\xi(\tau - \lambda) d\lambda \quad (3)$$

La varianza del proceso a la salida se halla asumiendo en (3),  $\tau = 0$ ,

$$\sigma_\eta^2 = R_\eta(0) = \frac{1}{2T} \int_{-2T}^{2T} \left(1 - \frac{|\lambda|}{2T}\right) R_\xi(\lambda) d\lambda$$

Entonces para el cálculo de la varianza del promedio deslizando en el tiempo de una señal estacionaria en el sentido amplio, es necesario conocer su función de correlación,

$$\sigma_\eta^2 = R_\eta(0) = \frac{1}{2T} \int_{-2T}^{2T} \left(1 - \frac{|\lambda|}{2T}\right) R_\xi(\lambda) d\lambda = \frac{\sigma_\xi^2}{2T} \int_{-2T}^{2T} \left(1 - \frac{\tau}{T}\right) r_\xi(\tau) d\tau \quad (4)$$

La expresión (4) presenta interés para los modelos de señales aleatorios, en la medida en que se analiza su valor asintótico,  $\lim_{T \rightarrow \infty} \sigma_\eta^2(T) = 0$ .

### 5.1.3. Empleo de operadores lineales

En la sección §3.1.1 se analiza la transformación de variables empleando operadores lineales. Un sistema lineal e invariante en el tiempo descrito por la ecuación diferencial (1.38), se puede representar empleando operadores en la siguiente forma:

$$\mathcal{A}\{\eta(t)\}(p, t) = \mathcal{B}\{\xi(t)\}(p, t) \quad (5.12)$$

donde  $\mathcal{A}(p, t) = \sum_m \alpha_m p^m$  y  $\mathcal{B}(p, t) = \sum_n \beta_n p^n$  son los operadores de salida y entrada respectivamente, siendo  $p = d/dt$ . De la expresión (5.12), se obtiene la relación formal entre la señal de entrada y salida, mediante los respectivos operadores:

$$\eta(t) = \frac{\mathcal{B}(p, t)}{\mathcal{A}(p, t)} \{\xi(t)\} = \mathcal{K}\{\xi(t)\}(p, t)$$



Empleando el operador  $\mathcal{K}(p, t)$  para la descripción del sistema, se pueden hallar las respectivas expresiones para los siguientes momentos estadísticos:

$$m_{1\eta}(t) = \mathcal{K}\{m_{1\xi}(t)\}(p, t) \quad (5.13a)$$

$$R_{\eta}(t_1, t_2) = \mathcal{K}\{\mathcal{K}\{R_{\xi}(t_1, t_2)\}(p, t_2)\}(p, t_1) \quad (5.13b)$$

**Ejemplo 5.7.** Sea un sistema diferenciador de primer orden, a cuya entrada se aplica la señal aleatoria  $\xi(t)$  con valor medio,  $m_{1\xi}(t) = \sin \omega t$ , y función de correlación de la forma  $R_{\xi}(t_1, t_2) = \sigma_{\xi}^2 \exp(-\alpha(t_2 - t_1))^2$ . Determinar la media y la varianza del proceso a la salida.

La señal aleatoria a la salida  $\eta(t)$  se relaciona con la entrada, mediante el operador lineal,  $\eta(t) = \mathcal{K}\{\xi(t)\}(p, t) = d\xi/dt$ , entonces, de las relaciones (5.13a) y (5.13b) se obtiene:

$$m_{1\eta}(t) = \frac{d}{dt} m_{1\xi}(t) = \omega \cos \omega t$$

$$R_{\eta}(t_1, t_2) = \frac{\partial^2}{\partial t_1 \partial t_2} R_{\xi}(t_1, t_2) = 2\sigma_{\xi}^2 \alpha e^{-\alpha(t_2 - t_1)^2} (1 - 2\alpha(t_2 - t_1)^2)$$

Haciendo  $t_2 = t_1 = t$  se encuentra  $R_{\eta}(t, t) = \sigma_{\eta}^2 = 2\alpha\sigma_{\xi}^2$ .

**Ejemplo 5.8.** Sea un sistema integrador,  $\eta(t) = \int_0^t \xi(\lambda) d\lambda$  a la entrada del cual se aplica una señal aleatoria  $\xi(t)$  con valor medio  $m_{1\xi}(t)$  y función de correlación  $R_{\xi}(\tau)$ . Determinar el valor medio y la función de correlación del proceso a la salida.

De manera similar al caso anterior, teniendo en cuenta que la integral corresponde a un operador lineal, se obtiene la relación entre los momentos de entrada y salida,

$$m_{1\eta}(t) = \int_0^t m_{1\xi}(t) dt$$

Haciendo  $m_{1\xi}(t) = m_{1\xi} = \text{const}$ , se observa que la media a la salida resulta variable en el tiempo,  $m_{1\eta}(t) = m_{1\xi}t$  (excepto para el caso  $m_{1\xi} = 0$ ), luego el proceso a la salida presenta un comportamiento no estacionario. A efecto de simplificar el análisis se asume  $m_{1\xi}(t) = 0$ , entonces la función de correlación, a partir de las relación (5.13b) resulta en:

$$R_{\eta}(t_1, t_2) = \int_0^{\lambda_1} \int_0^{\lambda_2} R_{\xi}(t_1, t_2) dt_1 dt_2$$

la cual muestra, que para integrar una señal aleatoria se exige la existencia de su función de correlación. Asumiendo además, la estacionariedad del proceso a la entrada, la función de correlación del proceso a la salida se resume hasta la integral doble

$$R_{\eta}(t_1, t_2) = \int_0^{\lambda_1} \int_0^{\lambda_2} R_{\xi}(t_1 - t_2) dt_1 dt_2$$

que depende por separado de  $t_2$  y  $t_1$  y no de su diferencia, por lo tanto, el proceso a la salida no es estacionario.

La varianza se puede obtener de la última integral, teniendo en cuenta (3.12b) [12]:

$$\sigma_{\eta}^2(t) = 2 \int_0^t (t - \tau) R_{\xi}(\tau) d\tau$$

En general, se puede decir que las FDP cambian a su paso por operadores lineales, excepto para el caso Gaussiano, que mantiene su estructura.

## Problemas

**Problema 5.1.** Se tiene una mezcla lineal,  $x(t) = s(t) + \eta(t)$ , ambos procesos con correlación mutua nula, donde  $s(t) = s_0 \cos(\omega t + \phi)$  es la señal útil y  $\eta(t)$  es el ruido con densidad normal  $\mathcal{N}(m_{1\eta}(t), \sigma_{\eta}(t))$ . Hallar la FDP de la suma.

**Problema 5.2.** Un proceso aleatorio  $\eta(t)$  está compuesto de una cantidad  $N \gg 1$  de señales aleatorias independientes  $\xi_i$ , cada una de ellas con FDP Rayleigh. Hallar la FDP del proceso suma.

**Problema 5.3.** Dados los procesos  $\xi_1(t) = a \cos \omega_c t$  y  $\xi_2(t) = b \sin \omega_c t$ , siendo  $\omega_c = \text{const}$ , y  $a, b$  variables aleatorias invariables en el tiempo. Hallar la esperanza, la función de correlación del proceso suma  $\eta(t) = \xi_1(t) + \xi_2(t)$  y establecer las condiciones para su estacionariedad.

**Problema 5.4.** Sea el proceso aleatorio  $\xi(t)$  con función de correlación  $R_{\xi}(\tau)$ , para el cual se define el proceso diferencial como  $\xi_{\Delta}(t) = \xi(t + \Delta t) - \xi(t)$  que contiene la información de la diferencia de valores del proceso aleatorio  $\xi(t)$  distanciados en el tiempo por el intervalo  $\Delta t$ . Demostrar las siguientes igualdades (donde  $R_{\Delta}(\tau)$  es la función de correlación del proceso diferencial):

$$R_{\Delta}(\tau) = 2R_{\xi}(\tau) - R_{\xi}(\tau + \Delta) - R_{\xi}(\tau - \Delta), \quad S_{\Delta}(\omega) = 4 \sin^2 \frac{\omega \Delta}{2} S_{\xi}(\omega)$$

**Problema 5.5.** Hallar la función de correlación mutua entre el proceso aleatorio  $\xi(t)$  y su derivada  $d\xi/dt$ , además hallar su relación con la respectiva DEP  $S_{\xi}(\omega)$ .

**Problema 5.6.** Sea la función de correlación de un proceso aleatorio  $\xi(t)$  de la forma  $R_{\xi}(\tau) = \sigma_{\xi}^2 \exp(-\alpha^2 \tau^2)$ . Hallar la función de correlación de la derivada del proceso  $d\xi/dt$  y calcular sus valores máximo y mínimo. Además hallar la función de correlación del proceso  $\eta(t) = \xi(t) + d\xi/dt$ .

**Problema 5.7.** Encontrar la diferencia en las funciones de correlación para el caso de RBG y ruido rosado a la salida a la salida del circuito  $RC$  (ejemplo 5.1).

**Problema 5.8.** Hallar la función de correlación de la integral del proceso aleatorio  $\xi(t)$  con función de correlación  $R_{\xi}(\tau) = \sigma_{\xi}^2 \exp(-\alpha |\tau|)$ .

**Problema 5.9.** Hallar la función de correlación y la respectiva DEP a la salida de cada uno de los circuitos del problema 1.12, cuando a la entrada se tiene RBG.

**Problema 5.10.** Hallar la función de correlación de salida del circuito mostrado en el ejemplo 5.1, si la función de correlación de entrada es la expresión  $R_{\xi}(\tau) = \frac{\alpha N_0}{2} \exp(-\alpha |\tau|)$

**Problema 5.11.** Sea el RBG a la entrada de un elevador al cuadrado ( $\eta(t) = \xi^2(t)$ ). Encontrar la FDP, la media y la varianza a la salida del sistema no lineal.

## 5.2. Paso de señales aleatorias por sistemas no lineales

Como en el caso de los sistemas lineales, el paso de procesos aleatorios por sistemas no lineales consiste en la determinación de las características de aleatoriedad de los procesos a la entrada y salida del sistema. Sin embargo, debido a la complejidad de solución en este caso, el análisis se enfoca en dos tareas principales: primero, dadas las funciones de no linealidad hallar la representación adecuada de la no linealidad de los sistemas y, segundo, determinar de forma aproximada los momentos y valores de aleatoriedad.

En la primera tarea, por cuanto no existen métodos regulares de solución de cualquier ecuación diferencial no lineal, pues tampoco se tienen métodos regulares de solución de ecuaciones diferenciales no lineales estocásticas. En este sentido, se ha desarrollado una serie de métodos de solución aproximada, entre los cuales están los siguientes:

### 5.2.1. Métodos de linealización

**Linealización determinística.** El método consiste en el cambio de la ecuación diferencial de tipo no lineal que describe el sistema por una (o varias) ecuaciones lineales diferenciales, dado algún criterio de equivalencia de representación. La condición principal del método exige que la no linealidad del sistema sea relativamente pequeña, lo que tiene lugar en el caso de análisis de procesos aleatorios de baja potencia (análisis de señal pequeña), tanto que las fluctuaciones en los incrementos del proceso aleatorio de salida sean menores que el valor estacionario de la reacción  $\eta_0$ , esto es,  $\sigma_\eta \ll \eta_0$ . El valor  $\eta_0$  se supone conocido, y usualmente corresponde al valor asintótico del parámetro en consideración (análisis de estado estable).

La ecuación de estacionariedad, denominada *ecuación de aproximación cero*, se define como la esperanza matemática de la función de no linealidad  $f$ , para un comportamiento medio de entrada

$$\mathcal{K}_0 \{ \eta_0 \} = E \{ f \{ \xi \} \}$$

donde  $\mathcal{K}_0$  es el funcional, que se obtiene del operador  $\mathcal{K}$  para la ecuación diferencial inicial  $\mathcal{K} \{ \eta(t) \} = f \{ \xi(t) \}$  con la condición de que todas sus derivadas sean iguales a cero. La ecuación de incremento diferencial expresada en términos de los incrementos de la salida, dados mediante la diferencia,  $\Delta\eta = \eta - \eta_0$ , se obtiene de la forma

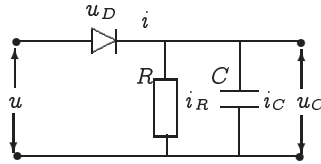
$$\mathcal{K}_0 \{ \eta_0 - \Delta\eta \} - \mathcal{K}_0 \{ \eta_0 \} = f \{ \xi \} - E \{ f \{ \xi \} \} \quad (5.14)$$

Seguidamente, se descompone la función no lineal con argumento  $\Delta\eta$  en la vecindad de  $\eta_0$  por alguna serie de potencia, conservando solo su primer término, de tal manera que se asegure la linealidad de la ecuación diferencial con respecto al argumento  $\Delta\eta$  (*aproximación cero*). La ecuación diferencial de aproximación obtenida sirve para el cálculo de los valores de aleatoriedad de los incrementos de  $\Delta\eta$ , de los cuales una vez obtenidos se calculan los correspondientes valores medios de la salida,  $\eta = \eta_0 + \Delta\eta$ .

**Ejemplo 5.9.** Dado el circuito del detector de amplitud, mostrado en la Figura 5.2, al cual le corresponde la ecuación diferencial

$$\dot{u}_c + \frac{1}{RC}u_c = \frac{1}{C}f(u - u_c),$$

cuyo elemento no lineal, el diodo, se aproxima por la función  $f(\lambda) = i_0 \exp(a\lambda)$   $a = \text{const.}$  Hallar la aproximación en forma de ecuación diferencial lineal, mediante el método de linealización determinística y su solución estacionaria para una entrada aleatoria  $u$  con FDP  $\mathcal{N}(0, \sigma_u)$ , asumiendo que  $\sigma_{u_c} \ll u_{C_0}$ .



**Figura 5.2.** Detector de amplitud

El método de linealización se desarrolla por los siguientes pasos:

1. Desarrollo del modelo a través de la diferencial no lineal. En este caso, la ecuación diferencial no lineal de partida, que tiene la forma

$$\dot{u}_c + \frac{1}{RC}u_c = \frac{1}{C}f(u - u_c) = \frac{1}{C}i_0 \exp(a(u - u_c)), \quad a = \text{const}$$

se puede replantear, de tal manera que en la parte derecha no esté presente la salida buscada,

$$\begin{aligned} \dot{u}_c \exp(au_c) + \frac{1}{RC}u_c \exp(au_c) &= \frac{i_0}{C} \exp(au) \\ \dot{v} + \frac{1}{RC}v \ln v &= \frac{ai_0}{C} e^{au}, \quad v = e^{au_c} \end{aligned} \quad (1)$$

2. Se halla la ecuación promediada de estacionariedad o ecuación de aproximación cero, haciendo en la anterior ecuación  $\dot{v}$  y promediando la parte derecha,

$$\frac{1}{RC}v_0 \ln v_0 = \frac{ai_0}{C} E\{e^{au}\}, \quad (2)$$

De la FDP  $p(u)$ , se halla FDP transformada  $p(v_1)$ , y luego se calcula el valor promedio  $E\{v_1\} = E\{\exp(au)\}$ . Sin embargo, en este caso, debido a la relación exponencial entre las variables, es preferible el empleo de la función característica. Haciendo el valor  $j\omega = a$  se obtiene

$$E\{e^{au}\} = E\{e^{j\omega u}\} = \theta_u(j\omega) = \exp\left(-\frac{\sigma_u^2}{2}\omega\right) = \exp\left(-\frac{\sigma_u^2 a^2}{2}\right)$$

El reemplazo de la última expresión en la ecuación (2) resulta en

$$v_0 \ln v_0 = aRi_0 \exp(\sigma_u^2 a^2 / 2) = au_{C_0} \exp(au_{C_0})$$

por cuanto,  $\exp(au_{C_0}) = v_0$ ,  $au_{C_0} = \ln v_0$ .

La ecuación obtenida se resuelve en términos de  $u_{C_0}$ , y consecuentemente, luego del cálculo de  $\sigma_{u_C}$  se comprueba el cumplimiento de la condición,  $\sigma_{u_C} \ll u_{C_0}$ .

3. Se halla la ecuación no lineal (5.14), combinando (1) con (2),

$$\dot{v} + \frac{1}{RC} v \ln v - \frac{1}{RC} v_0 \ln v_0 = \frac{a i_0}{C} (e^{au} - E\{e^{au}\})$$

de la cual se obtiene la respectiva ecuación para el incremento  $\Delta v = z$ .

Asumiendo que  $v = v_0 + \Delta v = v_0 + z$ , se obtiene

$$RC \dot{z} + (v_0 + z) \ln(v_0 + z) - v_0 \ln v_0 = a R i_0 (e^{au} - E\{e^{au}\}) \quad (3)$$

La ecuación no lineal  $v \ln v = (v_0 + z) \ln(v_0 + z)$  se descompone mediante la serie de Taylor en potencias de  $z$ ,

$$v \ln v = v_0 \ln v_0 + (1 + \ln v_0) z + \frac{1}{2v_0} z^2 - \dots$$

con lo cual, la ecuación de incremento (3) toma la forma,

$$RC \dot{z} + (1 + \ln v_0) z + \frac{1}{2v_0} z^2 - \dots = a R i_0 (e^{au} - E\{e^{au}\})$$

4. Se linealiza la ecuación de incremento, limitando su representación hasta el primer término:

$$\dot{z} + \frac{\ln v_0 + 1}{RC} z = \frac{a}{C} i_0 (e^{au} - E\{e^{au}\})$$

5. Se resuelve el modelo de ecuación diferencial lineal obtenido. Así, por cuanto,

$$\frac{dy}{dx} + Py = Q(x)$$

entonces, la integral general se halla de la expresión

$$y = \exp \left\{ - \int P dx \right\} \left( \int Q(x) \exp \left( \int P dx \right) dx + c \right), \quad P = \frac{\ln v_0 + 1}{RC}$$

Asumiendo la estacionariedad del proceso, entonces, se obtiene el valor asintótico para

$$t \rightarrow \infty, \quad c \exp \left( \int P dt \right) = c \exp(Pt) \rightarrow 0$$

Se debe tener en cuenta que la integración se realiza en el intervalo  $(-\infty, t)$ ,

$$z = e^{-Pt} \int_{-\infty}^t Q(\lambda) e^{P\lambda} d\lambda = \int_{-\infty}^t Q(\lambda) e^{-P(t-\lambda)} d\lambda$$

6. Se hallan las características de aleatoriedad del incremento  $\Delta v$ , de acuerdo a lo expuesto en el numeral §5.1.

**Linealización estadística.** El método, como en el caso anterior de linealización, reemplaza la transformación no lineal por una lineal, dados los criterios de equivalencia probabilística de la solución de ambas ecuaciones.

Sea la señal de salida  $\zeta(t)$  de un elemento con dependencia no lineal invariable en el tiempo,  $\zeta(t) = f(\eta(t))$ , la cual se propone reemplazar mediante la dependencia lineal  $\hat{\zeta} = k_0 + k\eta$ , siendo  $k$  el coeficiente equivalente de ganancia. Escogiendo adecuadamente un criterio de linealización se pueden determinar los coeficientes  $k, k_0$ , así por ejemplo, al tomar el criterio del mínimo error cuadrático medio (2.64) se tiene,

$$\varepsilon^2 = \min_{k_0, k} E \left\{ (\zeta - \hat{\zeta})^2 \right\} = E \left\{ (f(\eta) - k_0 - k\eta)^2 \right\}$$

con lo cual resulta el siguiente sistema de ecuaciones algebraicas,

$$\begin{aligned} \frac{\partial \varepsilon^2}{\partial k^2} &= 2(km_{2\eta} + k_0m_{1\eta} - E\{\zeta\eta\}) = 0, \\ \frac{\partial \varepsilon^2}{\partial k_0^2} &= 2(km_{1\eta} + k_0 - m_{1\zeta}) = 0 \end{aligned}$$

por lo tanto,

$$k = E\{(f(\eta) - m_{1\zeta})(\eta - m_{1\eta})\} / \sigma_\eta^2, \quad k_0 = m_{1\zeta} - km_\eta \quad (5.15)$$

La determinación de los valores de las respectivas constantes  $k_0, k$  en la expresión (5.15) implica el conocimiento de las funciones de los momentos de aleatoriedad del proceso de salida  $\eta(t)$ , lo cual en la práctica es parte de la tarea a calcular y limitando así el empleo de este método. Frecuentemente, una forma de solución aproximada está en asumir la FDP Gaussiana del proceso  $\eta(t)$ , que de manera evidente limita el alcance de la solución.

Una simplificación del método de linealización estadística consiste en el reemplazo de la ecuación diferencial no lineal por una lineal, cambiando los dispositivos de transformación variable en el tiempo por un equivalente invariable.

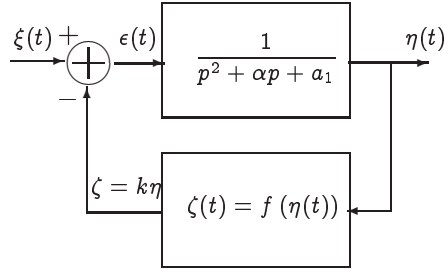
El criterio de equivalencia supone la invariabilidad de los valores de aleatoriedad del proceso en el tiempo, en particular se asume la restricción,

$$\tau_\xi \gg \tau_s \quad (5.16)$$

donde  $\tau_\xi$  es el tiempo de correlación de la señal de entrada y  $\tau_s$  es la constante de tiempo del sistema. Cabe anotar, que la condición (5.16) implica que no necesariamente se debe tener en cuenta el término asociado a  $\dot{\eta}$  con respecto a  $\eta$ .

---

**Ejemplo 5.10.** Sea el sistema no lineal de segundo orden, mostrado en la Figura 5.3, con señal de entrada  $\xi(t)$ , que corresponde al RBG con  $m_{1\xi} = 0$ . Hallar los valores de la media y la varianza del proceso  $\eta(t)$  a la salida.

**Figura 5.3.** Sistema lineal de segundo orden

Por cuanto  $m_{1\xi} = 0$ , entonces,  $m_{1\eta} = m_{1\zeta} = 0$  y la expresión (5.15) se simplifica:

$$k = E\{\eta(t) f(\eta(t))\} / \sigma_\eta^2, \quad k_0 = 0$$

Si se asume la FDP Gaussiana del proceso  $\eta(t)$ ,  $\mathcal{N}(0, \sigma_\eta)$ , además, para efectos de simplicidad, si la no linealidad se restringe a ordenes impares, esto es,  $\zeta = \sum_{i=2}^n a_{2i-1} \eta^{2i-1}$ , entonces la ganancia equivalente es:

$$k = E\{\eta f(\eta)\} / \sigma_\eta^2 = \sum_{i=2}^n a_{2i-1} E\{\eta^{2i}\} / \sigma_\eta^2$$

De otra parte, teniendo en cuenta la expresión para los momentos pares de la FDP normal, entonces,

$$k = 1 \cdot 3 a_3 \sigma_\eta^2 + \dots + 1 \cdot 3 \dots (2n-1) a_{2n-1} (\sigma_\eta^2)^{n-1}$$

El sistema linealizado de la Figura 5.3 tiene función de transferencia,

$$H(\omega) = 1 / \left( (j\omega)^2 + \alpha j\omega + a_1 + k \right)$$

que asumiendo la estacionariedad del proceso  $\eta(t)$ , implica la DEP,  $S_\eta(\omega) = (N_0/2) |H(\omega)|^2$ , de la cual se puede hallar el respectivo valor de varianza,

$$\begin{aligned} \sigma_\eta^2 &= \frac{1}{2\pi} \int_{-\infty}^{\infty} S_\eta(\omega) d\omega = \frac{N_0}{4\pi} \int_{-\infty}^{\infty} |H(\omega)|^2 d\omega \\ &= \frac{N_0}{4\alpha(a_1 + k)} \end{aligned}$$

**Ejemplo 5.11.** Sea el detector de amplitud representado en la Figura 5.2, en cuya entrada se tiene la señal  $u = u_0 \cos(\omega_0 t + \phi(t))$ , que cambia muy lentamente en el tiempo (señal de banda angosta), tal que tiene se puede asumir,  $RC \ll \tau_u$ ,  $RC \gg 2\pi/\omega_0$ , donde  $\tau_u$  es el tiempo de correlación de la envolvente. Hallar el modelo de transformación invariable y equivalente, dado con relación al voltage de salida.

Teniendo en cuenta las siguientes expresiones que describen el sistema,

$$i = i_R + i_C, \quad i_C = C \frac{du_C}{dt}, \quad i_R = \frac{1}{R} u_C, \quad u_D = u - u_C, \quad i = f(u_d)$$

entonces, la ecuación diferencial inicial con términos variables en el tiempo tiene la forma,

$$\begin{aligned} f(u - u_C) &= \frac{1}{R} u_C + C \frac{du_C}{dt} \\ \frac{du_C}{dt} + \frac{1}{RC} u_C &= \frac{1}{C} f(u - u_C) \end{aligned}$$

El modelo de transformación invariable en el tiempo exige analizar el término de la derivada para su eliminación. La condición de un tiempo de correlación de la envolvente de entrada mucho mayor que la inercia del sistema, para el caso de análisis de una señal senoidal con frecuencia  $\Omega$ , permite tomar la siguiente aproximación,

$$\Omega \approx \frac{1}{\tau_u} \quad (1)$$

Sea  $u_C = U_{C_0} \sin \Omega t$ , entonces,  $du_C/dt = \Omega u_{C_0} \cos \Omega t$ . Al tomar la aproximación (1), se tiene

$$\frac{u_{C_0}}{\tau_u} \cos \Omega t + \frac{u_{C_0}}{RC} \sin \Omega t = \frac{1}{C} f(u - u_C)$$

Teniendo en cuenta la condición  $\tau_u \gg RC$ , se observa que el primer término de la ecuación se puede eliminar, por lo tanto, la ecuación diferencial inicial se simplifica hasta el modelo

$$u_c = RC f(u - u_C)$$

que corresponde a una transformación invariable.

Cabe anotar que la otra desigualdad,  $RC \gg 2\pi/\omega_0$ , implica que tiene lugar la filtración de todas las componentes espectrales de alta frecuencia, por lo que se puede asumir la similitud de los anchos de banda de los espectros de las envolventes de entrada y salida.

**Transformación en modelos de Markov.** En aquellos casos en los cuales la salida se puede asumir como un proceso del Markoviano, descrito en el numeral §4.1, la FDP de salida  $p(\eta)$  se puede hallar empleando la ecuación de Fokker-Plank-Kolmogorov (4.34).

La presunción Markoviana se puede asumir si existe la independencia de los incrementos del proceso aleatorio  $\Delta\xi = \xi(t + \tau) - \xi(t)$  para intervalos separados de tiempo. No obstante, esta igualdad no es suficiente. Un criterio cualitativo al respecto puede ser un valor muy pequeño del tiempo de correlación  $\tau_\xi$  de la entrada con respecto a todas las constantes de tiempo del sistema.

Sea un sistema descrito por la más simple ecuación diferencial,  $\dot{\eta} = F(\xi, \eta)$ , dada la entrada  $\eta(t)$  con valor medio 0, entonces, la condición Markoviana (4.20) se asume si tiene lugar la desigualdad

$$\tau_\xi \ll \left( \frac{\partial F}{\partial \xi} \right)^{-1}$$

El análisis de sistemas de mayor orden implica su representación alterna en forma de ecuación vectorial diferencial de primer orden, como se muestra en el caso lineal en (1.61), para el cual se comprueban las respectivas desigualdades.



### 5.2.2. Series de Volterra

Dadas la señal aleatoria de entrada  $\xi(t)$  y salida  $\eta(t)$ , además, asumiendo que de alguna manera se puede determinar los núcleos y, por lo tanto, se puede representar el sistema no lineal con memoria finita por la serie de Volterra (1.44) truncada hasta  $n$  términos,

$$\eta(t) = h_0 + \sum_{k=1}^n \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} h_k(\tau_1, \dots, \tau_k) \xi(t - \tau_1) \cdots \xi(t - \tau_k) d\tau_1 \dots d\tau_k \quad (5.17)$$

entonces, la anterior representación permite analizar el sistema por el método de momentos, por el cual se hallan los momentos de la salida  $\eta(t)$  a partir del conocimiento de los momentos de la señal de entrada  $\xi(t)$ . En este caso, el cálculo de los momentos y valores de aleatoriedad se realiza de manera similar al caso de los sistemas lineales [15].

Así por ejemplo, empleando la operación de promedio a ambos lados de (5.17) se tiene,

$$\overline{\eta(t)} = E\{\eta(t)\} = h_0 + \sum_{k=1}^n \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} h_k(\lambda_1, \dots, \lambda_k) E\{\xi(t - \lambda_1) \cdots \xi(t - \lambda_k)\} d\lambda_1 \dots d\lambda_k$$

Así mismo, para el caso del segundo momento inicial, se obtiene,

$$\begin{aligned} \overline{\eta^2(t)} &= h_0^2 + 2h_0 \overline{\eta(t)} \\ &+ E \left\{ \left( \sum_{k=1}^n \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} h_k(\lambda_1, \dots, \lambda_k) \xi(t - \lambda_1) \cdots \xi(t - \lambda_k) d\lambda_1 \dots d\lambda_k \right)^2 \right\} \end{aligned}$$

De lo anterior, se observa que a medida que crece el orden de aproximación  $n$ , significativamente aumenta la dificultad de cálculo, motivo por el cual, el método de representación con series de Volterra, en la práctica, no se considera para valores  $n \gg 1$ .

**Ejemplo 5.12.** Resolver el ejemplo 5.10 mediante las series de Volterra.

El proceso de salida, formalmente se representa por la convolución  $\eta(t) = h(t) * \epsilon(t)$ , siendo

$$\epsilon(t) = \xi(t) - (a_3 \eta^3(t) + \cdots + a_{2n-1} \eta^{2n-1}(t))$$

mientras,  $h(t)$  es la respuesta a impulso de la parte lineal del sistema con función de transferencia,  $H(\omega) = (p^2 + \alpha p + a_1)^{-1}$ . Por lo tanto, se tiene,

$$\begin{aligned} \eta(t) &= \int_{-\infty}^{\infty} h(\tau) \xi(t - \tau) d\tau - a_3 \int_{-\infty}^{\infty} h(\tau) \eta^3(t - \tau) d\tau - a_1 \int_{-\infty}^{\infty} h(\tau) \eta^5(t - \tau) d\tau - \cdots \\ &\quad - a_{2n-1} \int_{-\infty}^{\infty} h(\tau) \eta^{2n-1}(t - \tau) d\tau \end{aligned}$$

Se debe hallar la representación del proceso a la salida en forma de la suma,

$$\eta(t) = \eta_1(t) + \eta_2(t) + \cdots + \eta_k(t) + \cdots$$

donde  $\eta_k(t)$  es la componente  $k$ , que corresponde al  $k$  núcleo. Al igualar ambas identidades anteriores se obtiene,

$$\begin{aligned} \sum_{k=1} \eta_k(t) &= \int_{-\infty}^{\infty} h(\tau) \xi(t-\tau) d\tau - a_3 \int_{-\infty}^{\infty} h(\tau) \eta^3(t-\tau) d\tau \\ &\quad - a_1 \int_{-\infty}^{\infty} h(\tau) \eta^5(t-\tau) d\tau - \cdots - a_{2n-1} \int_{-\infty}^{\infty} h(\tau) \eta^{2n-1}(t-\tau) d\tau \\ &= \int_{-\infty}^{\infty} h(\tau) \xi(t-\tau) d\tau - a_3 \int_{-\infty}^{\infty} h(\tau) \eta_1^3(t-\tau) d\tau \\ &\quad - 3a_3 \int_{-\infty}^{\infty} h(\tau) \eta_1^2(t-\tau) \eta_2(t-\tau) d\tau - 3a_3 \int_{-\infty}^{\infty} h(\tau) \eta_1(t-\tau) \eta_2^2(t-\tau) d\tau - \cdots \end{aligned}$$

Seguidamente, se pueden igualar los términos de igual potencia de  $\xi(t)$ , con lo cual se obtiene el siguiente sistema de expresiones recurrentes:

$$\begin{aligned} \eta_1(t) - \int_{-\infty}^{\infty} h(\tau) \xi(t-\tau) d\tau &= 0, \quad \eta_2(t) = 0, \\ \eta_3(t) + a_3 \int_{-\infty}^{\infty} h(\tau) \eta_1^3(t-\tau) d\tau &= 0, \quad \eta_4(t) = 0, \\ \eta_5(t) + 3a_3 \int_{-\infty}^{\infty} h(\tau) \eta_1^2(t-\tau) \eta_3(t-\tau) d\tau + a_5 \int_{-\infty}^{\infty} h(\tau) \eta_1^5(t-\tau) d\tau &= 0, \quad \cdots \end{aligned}$$

Del primer término, se puede concluir que  $h_1(\tau) = h(\tau)$ , que al reemplazar en el tercer término resulta en

$$\begin{aligned} \eta_3(t) &= -a_3 \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h(\tau) h(\lambda_1) h(\lambda_2) h(\lambda_3) \xi(t-\tau-\lambda_1) \xi(t-\tau-\lambda_2) \\ &\quad \times \xi(t-\tau-\lambda_3) d\lambda_1 d\lambda_2 d\lambda_3 d\tau = 0, \end{aligned}$$

Se puede demostrar que la función de transferencia de dimensión múltiple del sistema con núcleo  $h_k(\lambda_1, \dots, \lambda_k)$ , que precede a un sistema lineal con función de transferencia  $H(p)$ , se determina por la relación

$$H(p_1 + p_2 + \cdots + p_k) H_k(p_1, p_2, \dots, p_k)$$

Si por el contrario, el sistema lineal se encuentra antes del sistema con núcleo  $h_k(\lambda_1, \dots, \lambda_k)$ , la función de transferencia del todo el sistema será:  $H_k(p_1, p_2, \dots, p_k) H(p_1) H(p_2) \cdots H(p_k)$ .

Teniendo en cuenta la anterior, el tercer término de la descomposición,  $\eta_3(t)$ , se puede analizar como la salida de un sistema, que corresponde a la unión en cascada de un elemento no lineal con núcleo  $h(\lambda_1) h(\lambda_2) h(\lambda_3)$ , y un sistema lineal con respuesta a impulso  $-a_3 h(\tau)$ . Por lo tanto, la función de transferencia de tal sistema será:

$$H_3(p_1, p_2, p_3) = -a_3 H(p_1 + p_2 + p_3) H(p_1) H(p_2) H(p_3)$$

De manera similar, se obtiene la función de transferencia para el quinto término:

$$H_5(p_1, \dots, p_5) = H(p_1 + \dots + p_5) H(p_1) \dots H(p_5) (3a_2^2 H(p_3 + p_4 + p_5) - a_5)$$

A partir de estos primeros términos, se aproxima la correspondiente serie de descomposición para la varianza  $\sigma_\eta^2$  del proceso de salida:

$$\sigma_\eta^2 = E \left\{ \left( \sum_{k=1}^{\infty} \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} h_k(\tau_1, \dots, \tau_k) \xi(t - \lambda_1) \dots \xi(t - \lambda_k) d\lambda_1 \dots d\lambda_k \right) \right. \\ \left. \left( \sum_{k=1}^{\infty} \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} h_k(\tau_1, \dots, \tau_k) \xi(t - \lambda_1) \dots \xi(t - \lambda_k) d\lambda_1 \dots d\lambda_k \right) \right\}$$

Por cuanto el operador de promedio se puede introducir en el integral, entonces,

$$\sigma_\eta^2 = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h_1(\tau) h_1(\lambda_1) E\{\xi(t - \tau) \xi(t - \lambda_1)\} d\tau d\lambda_1 \\ + 2 \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h_1(\tau) h_3(\lambda_1, \lambda_2, \lambda_3) \\ \times E\{\xi(t - \tau) \xi(t - \lambda_1) \xi(t - \lambda_2) \xi(t - \lambda_3)\} d\tau d\lambda_1 d\lambda_2 d\lambda_3 + \dots$$

El proceso a la entrada  $\xi(t)$  es ruido blanco Gaussiano con media 0 y función de correlación  $R_\xi(t_m, t_n) = E\{\xi(t_m) \xi(t_n)\} = (N_0/2) \delta(t_m - t_n)$ , por lo tanto,

$$E\{\xi(t_1) \dots \xi(t_{2n+1})\} = 0, \quad E\{\xi(t_1) \dots \xi(t_{2n})\} = \sum \prod R_\xi(t_j, t_p) R_\xi(t_p, t_m) R_\xi(t_q, t_r)$$

donde la suma se lleva a cabo por todas las particiones de las diferentes parejas, mientras el producto se realiza por todas las parejas de elementos en cada partición.

De esta manera, se tiene que la varianza buscada es igual a,

$$\sigma_\eta^2 = \frac{N_0}{2} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h_1(\tau_1) h_1(\lambda_1) \delta(\lambda_1 - \tau) d\lambda_1 d\tau + \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h_1(\tau) h_3(\lambda_1, \lambda_2, \lambda_3) \\ \times (\delta(\lambda_1 - \tau) \delta(\lambda_2 - \lambda_3) + \delta(\lambda_2 - \tau) \delta(\lambda_3 - \lambda_1) + \delta(\lambda_3 - \tau) \delta(\lambda_1 - \lambda_2)) d\tau d\lambda_1 d\lambda_2 d\lambda_3$$

Sin embargo, el cálculo del anterior integral se simplifica empleando la transformada de Laplace de dimensión múltiple. Así, conocido que,

$$L\{y_i(t_1, \dots, t_i) y_k(t_{i+1}, \dots, t_{i+k})\} = Y_i(p_1, \dots, p_i) Y_k(p_{i+1}, \dots, p_{i+k})$$

donde

$$\begin{aligned} L\{y_k(t_1, \dots, t_k)\} &= Y_k(p_1, \dots, p_k) \\ &= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} y_k(t_1, \dots, t_k) \exp(-p_1 t_1 - \cdots - p_k t_k) dt_1 \dots dt_k \end{aligned}$$

luego,

$$\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} y_i(t_1, \dots, t_i) y_k(t_{i+1}, \dots, t_{i+k}) dt_1 \dots dt_{i+k} = Y_i(0, \dots, 0) Y_k(0, \dots, 0)$$

El cálculo del primer término de la varianza se realiza de la siguiente forma:

$$L\{h_1^2(\lambda)\} = L\{h_1(\lambda_1) h_1(\lambda_2)\}_{p_1 \Leftrightarrow p_2} = \frac{1}{(p_1^2 + \alpha p_1 + a_1)(p_2^2 + \alpha p_2 + a_1)} \Big|_{p_1 \Leftrightarrow p_2}$$

que por tabla se llega a la siguiente expresión:

$$L\{h_1^2(\lambda)\} = \frac{2}{(p_1 + \alpha)(p_1^2 + 2\alpha p_1 + 4a_1)},$$

luego,

$$\int_{-\infty}^{\infty} h_1^2(\lambda) d\lambda = \frac{1}{2\alpha a_1}$$

En el cálculo del segundo término de la varianza, se realiza de forma similar al anterior,

$$L\{h_1(\lambda_1) h_3(\lambda_2, \lambda_3, \lambda_3)\} = -a_3 H(p_1) H(p_2) H(p_2 + p_3) \left( H(p_3) H(p_4) \Big|_{p_3 \Leftrightarrow p_4} \right)$$

Por lo tanto,

$$\int_{-\infty}^{\infty} h_1(\lambda_1) h_3(\lambda_2, \lambda_3, \lambda_3) d\lambda_3 = -\frac{a_3 H(p_1) H^2(p_2)}{2\alpha a_1}$$

De forma similar, se tiene que

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h_1(\lambda_1) h_3(\lambda_1, \lambda_3, \lambda_3) d\lambda_1 d\lambda_3 = -\frac{a_3}{2\alpha a_1} \left( H(p_1) H^2(p_2) \Big|_{p_1 \Leftrightarrow p_2} \right) \Big|_{p_1=0}$$

Al emplear el método de asociación de variables, se obtiene el valor

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h_1(\lambda_1) h_3(\lambda_1, \lambda_3, \lambda_3) d\lambda_1 d\lambda_3 = -\frac{a_3}{2\alpha a_1} \frac{1}{4\alpha a_1^2}$$

Finalmente, como resultado de todos los cálculos, se obtienen los siguientes dos primeros términos de la descomposición en serie de la varianza del proceso a la salida,

$$\sigma_\eta^2 = N_0 / 4\alpha a_1 - 3(a_3/a_1)(N_0/4\alpha a_1)^2 + \dots$$

que en forma general, converge con la expresión obtenida en el ejemplo 5.10.

## Problemas

**Problema 5.12.** Sea una señal de voltaje aleatoria con FDP  $\mathcal{N}(0, \sigma_v)$ , que alimenta el elemento no lineal con función de paso,

$$i(v) = \begin{cases} i_0 - 2vi_0/v_c + v^2i_0/v_c^2, & v \geq v_c \\ 0, & v < v_c \end{cases}$$

Asumiendo que la potencia del proceso es suficientemente pequeña, tal que se cumple la relación,  $|v_c| > 3\sigma_v$ . Hallar la FDP de la corriente a la salida.

**Problema 5.13.** La señal de voltaje aleatoria con FDP  $\mathcal{N}(0, \sigma_v)$  alimenta el elemento no lineal,  $i(v) = Svu(v)$ . Hallar el valor medio y la varianza de la corriente.

**Problema 5.14.** Hallar la función de correlación  $R_i(\tau)$  de la señal de corriente que pasa por un elemento lineal del tipo,  $i(v) = Svu(v)$ , cuando a la entrada del sistema se tiene un proceso estacionario  $\mathcal{N}(0, \sigma_v)$  y función de correlación  $r(\tau)$ .

**Problema 5.15.** Hallar la función de correlación  $R_i(\tau)$ , el valor medio y la varianza de una señal de corriente a la salida de un elemento con característica de paso aproximada por el polinomio  $i = a_0 + a_1v + a_2v^2$ , cuando a la entrada se tiene un proceso estacionario normal  $\mathcal{N}(0, \sigma_v)$ .

**Problema 5.16.** Hallar la función de correlación  $R_y(\tau)$ , del proceso a la salida del elemento con característica

$$y(x) = \begin{cases} -a, & x < 0 \\ a, & x \geq 0 \end{cases}$$

Asumir que  $R_x(0) = 1$ .



### 5.3. Transformación de procesos de Markov

#### 5.3.1. Transformación de señales aleatorias reales por procesos de Markov

En la práctica, el análisis de procesos aleatorios mediante los modelos de Markov exige que se constante la correspondiente condición markoviana, así como la determinación de los coeficientes respectivos de las ecuaciones de Kolmogorov. Aunque en forma estricta, los procesos reales no son del tipo Markoviano para una dimensión, en la medida en que los procesos de Markov no deben cumplir la condición de diferenciabilidad, mientras los reales si presentan esta propiedad, derivada de la existencia obligatoria de inercia en cualquier sistema dinámico físicamente implementable.

En este sentido, para generar los modelos de aproximación de señales reales empleado procesos de Markov con dimensión múltiple, se considera suficiente asumir la estacionariedad de la señal aleatoria, que tenga FDP Gaussiana, con DEP  $S_x(\omega)$  descrita por una función racional quebrada,

$$S_x(\omega) = \frac{|P_m(j\omega)|^2}{|Q_n(j\omega)|^2}, \quad m < n, S_x(\omega) \in \mathbb{Q} \quad (5.18)$$

donde  $Q_n(x) = x^n + \alpha_1 x^{n-1} + \dots + \alpha_n$  y  $P_m(x) = \beta_0 x^m + \beta_1 x^{m-1} + \dots + \beta_m$ , que corresponden al modelo de solución estacionario de la ecuación diferencial

$$\frac{d^n x}{dt^n} + \alpha_1 \frac{d^{n-1} x}{dt^{n-1}} + \dots + \alpha_n x(t) = \beta_0 \frac{d^m \xi}{dt^m} + \dots + \beta_m \xi(t) \quad (5.19)$$

donde  $\xi(t)$  es RBG con función de correlación,  $R_\xi(\tau) = \delta(\tau)$ .

En general, una ecuación diferencial de orden mayor se puede representar en forma de ecuación diferencial de primer orden, como se explica en (1.61), así,

$$\begin{aligned} \frac{dx_n}{dt} + \alpha_n x_1 + \alpha_{n-1} x_2 + \dots + \alpha_1 x_n + c_{n-m} \xi^{(m)}(t) + (c_{n-m+1} + \alpha_1 c_{n-m}) \xi^{(m-1)}(t) + \dots \\ \dots + (c_{n-1} + \alpha_1 c_{n-2} + \dots + \alpha_{m-1} c_{n-m}) \xi^{(1)}(t) + (\alpha_1 c_{n-1} + \alpha_2 c_{n-2} + \dots + \alpha_m c_{n-m}) \xi(t) \\ = \beta_0 \xi^{(m)}(t) + \beta_1 \xi^{(m-1)}(t) + \dots + \beta_{m-1} \xi'(t) + \beta_m \xi(t) \end{aligned} \quad (5.20)$$

donde las constantes  $c_{n-m}, c_{n-m+1}, \dots, c_{n-1}$  se escogen, de tal forma, que al eliminar en (5.19) las derivadas de mayor orden de la función  $x(t)$ , mediante el reemplazo por  $x_k(t)$ , se obtenga una ecuación diferencial de primer orden pero sin términos que contengan derivadas de  $\xi(t)$ .

Al relacionar los coeficientes con igual valor de derivada de  $\xi(t)$  a ambos lados de la ecuación, se obtiene la siguiente expresión recurrente para el cálculo de los coeficientes  $c_{kj}$ :

$$c_k = \beta_{k+m-n} - \sum_{j=1}^{k+m-n} a_j c_{k-j}, \quad k = n-m, n-m+1, \dots, n$$

con lo cual, la ecuación (5.20) se puede escribir en la forma

$$\frac{dx_n}{dt} = - \sum_{j=1}^n \alpha_{n+1-j} x_j + c_n \xi(t) \quad (5.21)$$

El sistema (5.21) contiene  $n$  ecuaciones de primer orden con las correspondientes  $n$  funciones desconocidas  $x, \dots, x_n$ . Además, la función  $\xi(t)$  se mantiene en la parte derecha de la igualdad. Debido a que el comportamiento del conjunto de funciones  $\{x_i\}$  se determina unívocamente por las condiciones iniciales de las mismas, entonces, estas son componentes de un proceso de Markov con dimensión  $n$ . Así mismo, la función inicial  $x(t)$ , por ser  $x(t) = x_i(t)$ , es una componente del proceso de Markov.

Los coeficientes del modelo de Markov,  $a_l$  y  $b_{lm}$ , se obtienen del sistema de ecuaciones (5.21) que tiene la misma estructura de representación de (4.25a), en la cual está presente una sola función  $\xi(t)$ , entonces se obtienen los coeficientes,

$$a_l = \begin{cases} y_{l+1}, & 1 \leq l \leq n-1 \\ - \sum_{j=1}^n \alpha_{n+1-j} y_j, & l = n \end{cases}$$

$$b_{lm} = \begin{cases} 0, & 1 \leq l \leq n-m-1 \\ c_l c_m, & n-m \leq l \leq n \end{cases}, (l \leq m)$$

La última expresión determina  $b_{lm}$ , para cualquiera que sean los índices, debido a la simetría de estos coeficientes.

La transformación a procesos de Markov, además del caso de un proceso normal estacionario con espectro racional quebrado, también es posible extenderlo al análisis de sistemas dinámicos, que contengan elementos (tanto lineales, como no lineales), a la entrada de los cuales estén presentes procesos normales con DEP fraccionales quebradas.

Sea el sistema dinámico descrito por la ecuación:

$$\frac{d^r z}{dt^r} + \gamma_1 \frac{d^{r-1} z}{dt^{r-1}} + \dots + \gamma_r z = f(x, t) \quad (5.22)$$

donde  $r_j$  son las constantes del sistema,  $f(x)$  es la función que describe un elemento no lineal invariante,  $x(t)$  es la señal aleatoria, de carácter estacionario y con estructura Gaussiana, que tiene densidad espectral de potencia en la forma (5.18). Reemplazando el proceso aleatorio  $x(t)$  por uno de Markov con dimensión  $n$ , de acuerdo a la expresión (1.61), y notando cada término como

$$x_{n+1}(t) = z(t), \quad x_{n+2}(t) = \dot{x}_{n+1}, \quad \dots, \quad x_{n+r} = \dot{x}_{n+r-1}(t) \quad (5.23)$$

entonces, la descripción del sistema (5.22) toma la forma:

$$\frac{dx_{n+r}}{dt} = -\gamma_1 x_{n+r} - \gamma_2 x_{n+r-1} - \dots - \gamma_r x_{n+1} + f(x_n, t) \quad (5.24)$$



Las expresiones (5.21), (5.23) y (5.24) conforman un sistema de  $(n + r)$  ecuaciones de primer orden, en la parte derecha de las cuales, además de la función  $x_j$ , está presente solo la señal aleatoria  $\xi(t)$  que tiene las propiedades del ruido blanco Gaussiano (en el sentido angosto). El sistema obtenido de ecuaciones corresponde a un proceso de Markov con dimensión  $n + r$ . En general, la transformación de señales reales a modelos de Markov es viable en los casos, en los que el proceso en análisis se describe por una ecuación diferencial, que puede ser no lineal, que contiene a la entrada una señal aleatoria,  $\xi(t)$ , con propiedades de RBG y que acepta su descomposición en potencias.

---

**Ejemplo 5.13.** Reemplazar el proceso estacionario normal  $x_1(t)$  con valor medio 0 y función de correlación  $R_{x_1}(\tau) = \sigma_x^2 e^{-\alpha|\tau|} \cos \beta \tau$ , por un proceso de Markov de 2 dimensiones.

La respectiva densidad espectral de potencia  $S_x(\omega)$  se determina como:

$$\begin{aligned} S_x(\omega) &= \frac{1}{2} \frac{\sigma^2}{2\pi} \left( \int_{-\infty}^{\infty} e^{-j\omega\tau - \alpha|\tau| + i\beta\tau} d\tau + \int_{-\infty}^{\infty} e^{-j\omega\tau - \alpha|\tau| + i\beta\tau} d\tau \right) \\ &= \frac{1}{2} \frac{\sigma^2}{2\pi} \left( \frac{\alpha}{(\omega - \beta)^2 + \alpha^2} + \frac{\alpha}{(\omega + \beta)^2 + \alpha^2} \right) = \frac{\sigma^2}{\pi} \frac{\omega^2 + \alpha^2 + \beta^2}{(\omega^2 - \beta^2 - \alpha^2)^2 + 4\alpha^2\omega^2} \\ &= \frac{\sigma^2}{\pi} \frac{|j\omega + \sqrt{\alpha^2 + \beta^2}|^2}{|-\omega^2 + 2j\alpha\omega + \alpha^2 + \beta^2|^2} \end{aligned}$$

Por lo tanto,  $x_1(t)$  es la solución estacionaria de la ecuación diferencial:

$$\frac{d^2 x_1}{dt^2} + 2\alpha \frac{dx_1}{dt} + (\alpha^2 + \beta^2)x_1 = \sqrt{2\alpha} \left( \frac{d\xi}{dt} + \sqrt{\alpha^2 + \beta^2}\xi \right) \sigma$$

donde  $\xi(t)$  es ruido blanco Gaussiano.

Cabe anotar que la señal aleatoria  $x_1(t)$  no es del tipo Markoviano, en la medida en que se determina por una ecuación de segundo orden, con lo cual el conocimiento de sus valores iniciales no es suficiente para la determinación de sus propiedades en el tiempo.

Sin embargo, en correspondencia con (1.61) haciendo  $\dot{x}_1(t) = x_2(t) + c_1\xi(t)$ , y realizando el respectivo reemplazo se obtiene:

$$\begin{cases} \frac{dx_2}{dt^2} + c_1 \frac{d\xi}{dt} + 2\alpha x_2 + 2\alpha c_1 \xi + (\alpha^2 + \beta^2)x_1 = \sqrt{2\alpha} \left( \frac{d\xi}{dt} + \sqrt{\alpha^2 + \beta^2}\xi \right) \sigma \\ \frac{d^2 x_1}{dt^2} + 2\alpha \frac{dx_1}{dt} + (\alpha^2 + \beta^2)x_1 = \sqrt{2\alpha} \left( \frac{d\xi}{dt} + \sqrt{\alpha^2 + \beta^2}\xi \right) \sigma \end{cases}$$

si se hace  $c_1 = \sigma\sqrt{2\alpha}$ , en la última expresión la derivada  $d\xi/dt$  desaparece y se obtiene un sistema de dos ecuaciones con dos variables desconocidas, que contienen RBG, y por lo tanto, que determinan el siguiente proceso de Markov con componentes  $x_1(t)$  y  $x_2(t)$ :

$$\begin{cases} \frac{dx_1(t)}{dt} - x_2(t) &= \sigma\sqrt{2\alpha}\xi(t) \\ \frac{dx_2(t)}{dt} + (\alpha^2 + \beta^2)x_1(t) + 2\alpha x_2(t) &= \sigma\sqrt{2\alpha} \left( \sqrt{\alpha^2 + \beta^2} - 2\alpha \right) \xi(t) \end{cases}$$


---

### 5.3.2. Modelos de transformación de señales

En los sistemas reales, las señales de información se pueden aproximar por un proceso aleatorio  $\mathbf{x}(t)$ , con DEP descrita por una función fraccional, (5.18):

$$S_x(\omega) = \frac{N_x}{2} \left| \frac{b_m(j\omega)^m + b_{m-1}(j\omega)^{m-1} + \cdots + b_0}{(j\omega)^n + a_{n-1}(j\omega)^{n-1} + \cdots + a_0} \right|^2 \quad (5.25)$$

en la cual, la condición de realización física del proceso  $\mathbf{x}(t)$  exige que  $m < n$ , esto es, que el proceso tenga potencia finita, además, las raíces del polinomio del denominador deben estar localizadas en la parte negativa real. De lo anteriormente expuesto, un proceso  $\mathbf{x}(t)$  se puede formar al pasar RBG  $\xi(t)$  con valor medio,  $m_{1\xi} = 0$ , y función de correlación  $R_\xi = \frac{N_\xi}{2}\delta(\tau)$ , a través de un filtro lineal con función de transferencia,

$$H_x(p) = \frac{b_m p^m + b_{m-1} p^{m-1} + \cdots + b_0}{p^n + a_{n-1} p^{n-1} + a_{n-2} p^{n-2} + \cdots + a_0} \quad (5.26)$$

donde  $p = d/dt$  es el operador de diferenciación.

En este caso, el proceso Gaussiano estacionario  $\mathbf{x}(t)$  con DEP (5.25), que se obtiene a la salida del filtro (5.26), se describe por la ecuación diferencial de  $n$  orden:

$$\frac{d^n x}{dt^n} + a_{n-1} \frac{d^{n-1} x}{dt^{n-1}} + \cdots + a_0 x(t) = b_m \frac{d^m \xi}{dt^m} + \cdots + b_0 \xi(t), \quad t \geq t_0 \quad (5.27)$$

La ecuación diferencial escalar (5.27) de orden  $n$  se puede reducir a la forma equivalente de una ecuación diferencial vectorial de primer orden,

$$\frac{d\mathbf{x}}{dt} = \mathbf{A}\mathbf{x}(t) + \mathbf{B}\xi(t), \quad \mathbf{x}(t_0) = \mathbf{x}_0, \quad t \geq t_0 \quad (5.28)$$

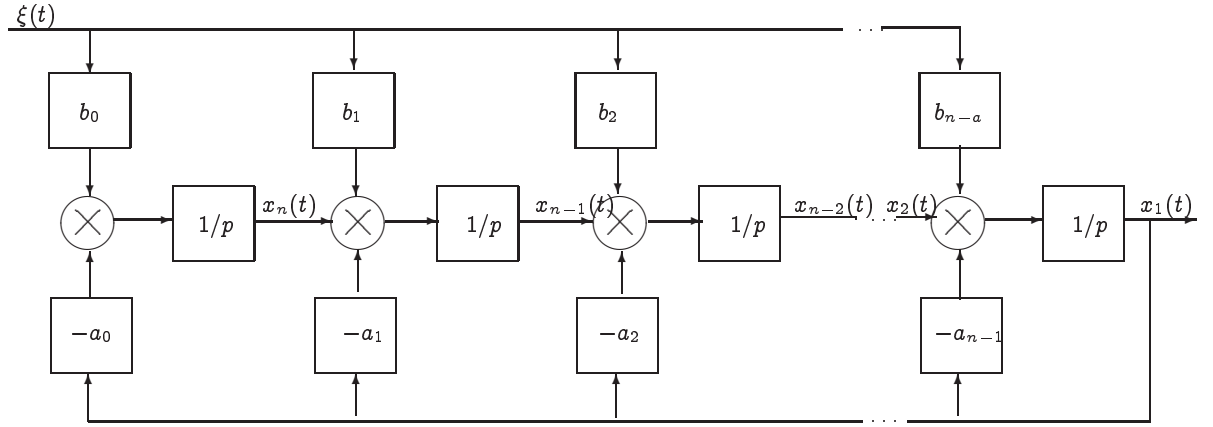
donde

$$\mathbf{A} = \begin{bmatrix} -a_{n-1} & 1 & 0 & 0 & \cdots & 0 \\ -a_{n-2} & 0 & 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & & \vdots \\ -a_0 & 0 & 0 & 0 & \cdots & 0 \end{bmatrix}, \quad \mathbf{x}(t) = \begin{bmatrix} x_1(t) \\ x_2(t) \\ \vdots \\ x_n(t) \end{bmatrix}, \quad \mathbf{B} = \begin{bmatrix} b_m \\ b_{m-1} \\ \vdots \\ b_0 \end{bmatrix} \quad (5.29)$$

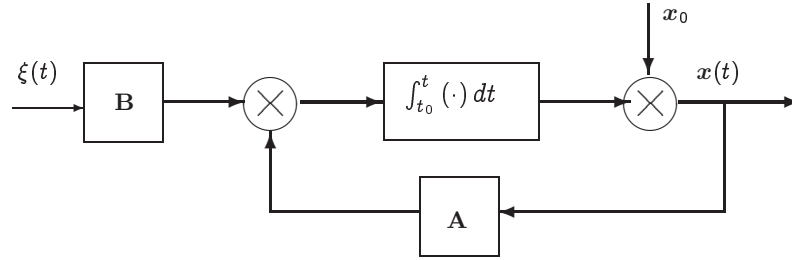
siendo  $\mathbf{x}_0$  el vector de condiciones iniciales, que corresponde a los valores del vector de estados  $\mathbf{x}(t)$  en el momento de tiempo  $t_0$ .

El modelo del filtro con respuesta de frecuencia (5.26) se muestra en la Figura 5.4(a), mientras en la Figura 5.4(b) se muestra el diagrama general del sistema lineal, descrito por la ecuación (5.28), en el cual los vectores se notan con una línea ancha. En forma general, las matrices  $\mathbf{A}(t)$  y  $\mathbf{B}(t)$ , tienen elementos que son funciones del tiempo y se determinan de los parámetros físicos del sistema dinámico.

La ecuación (5.28) describe un proceso Gaussiano de Markov, cuya realización corresponde a la función vectorial  $\mathbf{x}(t)$ , por lo tanto, cualquier proceso, sea estacionario o no,



(a) Filtro formador de proceso de Markov



(b) Diagrama general del sistema de formación

**Figura 5.4.** Modelo formación de procesos de Markov sobre sistemas lineales

pero con DEP fraccional y varianza finita, se puede realizar al pasar RBG por un sistema lineal descrito por la función de transferencia adecuada.

En general, las condiciones iniciales  $\mathbf{x}(t_0)$  en (5.28) son valores aleatorios, con FDP  $p(\mathbf{x}_0)$  estadísticamente independientes de la señal aleatoria  $\xi(t)$ . En principio, el conocimiento de  $p(\mathbf{x}_0)$  es suficiente para la descripción de la ecuación diferencial vectorial (5.28), aunque, también es suficiente la descripción del proceso aleatorio en el intervalo de tiempo  $(t_0, t)$  por la FDP con dimensión  $m$ , tal que  $m \rightarrow \infty$ , esto es, por el funcional de distribución,

$$F\{\mathbf{x}(t)\} = \lim_{\substack{T \rightarrow 0 \\ m \rightarrow \infty}} p_{t_0, t_1, \dots, t_n}(\mathbf{x}_0, \mathbf{x}_1, \dots, \mathbf{x}_m) \quad (5.30)$$

donde  $T = t_{t+\tau} - t_i$ ,  $i = 0, 1, \dots, m$ .

El funcional de distribución (5.30) depende de cada observación del vector aleatorio  $\mathbf{x}(t)$  en la medida en que la FDP  $p_{t_0, t_1, \dots, t_n}(\mathbf{x}_0, \mathbf{x}_1, \dots, \mathbf{x}_m)$  depende de los valores, que toman las variables aleatorias  $\mathbf{x}_0, \dots, \mathbf{x}_m$ .

Los funcionales de FDP existen para una clase limitada de procesos aleatorios, en particular, los Gaussianos y los de Markov. En el caso de los procesos de Markov, la FDP  $p_{t_0, \dots, t_m}(\mathbf{x}_0, \dots, \mathbf{x}_m)$  se puede expresar mediante la densidad de probabilidad inicial  $p_{t_0}(\mathbf{x}_0)$

y la FDP de transición del estado  $\mathbf{x}_{i-1}$  en el momento  $t_{i-1}$  al estado  $\mathbf{x}_i$  en el momento  $t_i$ ,

$$p_{t_i, t_{i+1}}(\mathbf{x}_i | \mathbf{x}_{i-1}) = \pi_{t_i, t_{i-1}}(\mathbf{x}_i | \mathbf{x}_{i-1}), \quad t_{i-1} < t_i, \quad i = 0, 1, \dots, m$$

esto es,

$$p_{t_0, t_1, \dots, t_m}(\mathbf{x}_0, \mathbf{x}_1, \dots, \mathbf{x}_m) = p_{t_0}(\mathbf{x}_0) \prod_{i=1}^m \pi_{t_{i-1}, t_i}(\mathbf{x}_{i-1}, \mathbf{x}_i) \quad (5.31)$$

De (5.31), la FDP con dimensión  $m+1$  para un proceso de Markov, toma la forma,

$$p_{t_0, t_1, \dots, t_m}(\mathbf{x}_0, \mathbf{x}_1, \dots, \mathbf{x}_m) = p_{t_0}(\mathbf{x}_0) \prod_{i=1}^{m+1} \pi_{t_{i-1}, t_i}(\mathbf{x}_{i-1}, \mathbf{x}_i) = \\ \pi_{t_m, t_{m+1}}(\mathbf{x}_{i-1}, \mathbf{x}_i) p_{t_0, t_1, \dots, t_m}(\mathbf{x}_0, \mathbf{x}_1, \dots, \mathbf{x}_m)$$

De lo anterior, se concluye que la FDP inicial más la FDP de transición  $\pi_{t_{i-1}, t_i}(\mathbf{x}_{i-1}, \mathbf{x}_i)$  determinan completamente el proceso vectorial aleatorio de Markov. Por cierto, ambas FDP cumplen, tanto las formas directa e inversa de la ecuación de Kolmogorov.

### 5.3.3. Ejemplos de procesos generados de Markov

En la práctica, se han generado modelos concretos que corresponden a señales reales, las cuales se analizan como procesos de Markov, descritos por (5.28) y por las respectivas FDP  $p_{t_0}(\mathbf{x}_0)$  y  $\pi_{t_{i-1}, t_i}(\mathbf{x}_{i-1}, \mathbf{x}_i)$ . Entre los modelos más conocidos, están los siguientes:

*Señal pasabaja de información.* Sea  $n = 1$ ,  $b_0 = a_0 = \alpha$ . Entonces de (5.28) se obtiene:

$$\frac{dx}{dt} = -\alpha x(t) + \alpha \xi(t), \quad x(t_0) = x_0 \quad (5.32)$$

El modelo (5.32) tiene la misma estructura del circuito  $RC$  presentado en el ejemplo 1.3, con constante de tiempo  $1/\alpha = RC$ , cuando a la entrada se tiene RBG,  $\xi(t)$ . En caso de no existir el primer término de (5.32), se obtiene un proceso estacionario de Wiener.

*Modelo con dos componentes de procesos de Markov de baja frecuencia.*

Sea  $n = 2$ . Entonces, se obtiene el sistema:

$$\begin{aligned} \frac{dx_1}{dt} &= x_2(t) - a_1 x_1(t) + b_1 \xi(t) \\ \frac{dx_2}{dt} &= -a_0 x_1(t) + b_0 \xi(t) \end{aligned} \quad (5.33)$$

En la práctica, es frecuente el empleo de modelos que incluyen dos componentes de procesos de Markov, que tienen una descripción diferente a (5.33), pero que pueden ser descritos por la ecuación diferencial de segundo orden,

$$\frac{d^2 x_1}{dt^2} = a_1 \frac{dx_1}{dt} + a_0 x_1(t) = b_0 \xi(t) + b_1 \frac{d\xi}{dt}$$

que se representa en forma de un sistema de dos ecuaciones lineales diferenciales:

$$\begin{aligned}\frac{dx_1}{dt} &= -\alpha_1 x_1(t) + \alpha_1^2 x_2(t) + \beta_1 \xi(t) \\ \frac{dx_2}{dt} &= -\alpha_2 x_2(t) + \beta_2 \xi(t)\end{aligned}\quad (5.34)$$

donde

$$\alpha_1 + \alpha_2 = a_1, \alpha_1 \alpha_2 = a_0, \alpha_2 \beta_1 = \alpha_1^2, \beta_2 = b_0, b_1 = b_1$$

En particular, cuando  $\alpha_1^2 = \beta_1 = \alpha_2 = \beta_2$ , el proceso aleatorio  $x_1(t)$ , se puede emplear en el modelado de la señal de voz [41]. En este caso,  $x_1(t)$ , se analiza como la salida de un par de circuitos  $RC$  conectados en cascada (sin tener en cuenta su reacción mutua), uno de ellos de baja frecuencia con constante de tiempo  $1/a_2 = R_1 C_1$ , y el otro de alta frecuencia, con  $1/\alpha_1' = R_2 C_2$ , cuando a la entrada de la conexión se tiene RBG. De otra parte, cuando se asumen los valores  $\alpha_1^1 = -\alpha_0'$ ,  $\beta_1 = 0$ ,  $\alpha_2 = \beta_2$ , de (5.34) se obtiene el modelo de trayectorias simples de seguimiento, en el cual el RBG pasa a través de dos circuitos de integración  $RC$  conectados en cascada, con las constantes de tiempo  $1/\alpha_1 = R_1 C_1$  y  $1/\alpha_2 = R_2 C_2$ , respectivamente. Cabe anotar que la conexión en cascada de circuitos  $RC$  se puede generalizar hasta cualquier orden  $n$ , en la medida en que el proceso  $x_1(t)$  se puede representar en la forma (5.34), para cualquiera que sea el orden de la ecuación diferencial (5.27) que lo describe:

$$\frac{dx}{dt} = \mathbf{A}x(t) + \mathbf{B}\xi(t), \quad x(t_0) = x_0, \quad t \geq t_0 \quad (5.35)$$

donde

$$\mathbf{A} = \begin{bmatrix} -\alpha_1 & -\alpha_1^2 & 0 & 0 & \cdots & 0 \\ 0 & -\alpha_2 & -\alpha_2^3 & 0 & \cdots & 0 \\ 0 & 0 & -\alpha_3 & -\alpha_3^4 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & 0 & \cdots & \alpha_n \end{bmatrix}, \quad \mathbf{B} = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_n \end{bmatrix}, \quad x(t) = \begin{bmatrix} x_1(t) \\ x_2(t) \\ \vdots \\ x_n(t) \end{bmatrix}$$

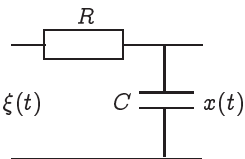
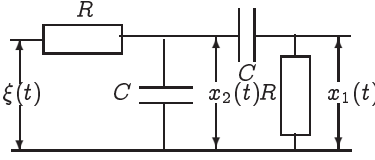
Los coeficientes en (5.27) se relacionan con los valores de (5.35) mediante las siguientes expresiones:

$$\begin{aligned}a_{n-1} &= \sum_{k=1}^n a_k; \quad a_{n-2} = \sum_{i=1}^{n-1} a_i \sum_{j=i+1}^n a_j \\ a_{n-3} &= \sum_{k=1}^{n-2} a_k \sum_{i=k+1}^{n-1} a_i \sum_{j=i+1}^n a_j; \quad a_{n-4} = \sum_{m=1}^{n-3} a_m \sum_{k=m+1}^{n-2} a_k \sum_{i=k+1}^{n-1} a_i \sum_{j=i+1}^n a_j; \\ a_0 &= a_1 a_2 a_3 \cdots a_{n-2} a_{n-1} a_n;\end{aligned}$$

De manera similar, se obtienen los coeficientes para la matriz  $\mathbf{B}$ ,

$$\begin{aligned}
 b_{n-1} &= \beta_1; b_{n-2} = \beta_1 \sum_{j=2}^n a_j - b_2 a_1^{(2)}; \\
 b_{n-3} &= \beta_1 \sum_{i=2}^{n-1} a_i \sum_{j=i+1}^n a_j - \beta_2 \alpha_1^{(2)} \sum_{k=3}^n a_k - \beta_3 \alpha_1^{(2)} a_2^{(3)}; \\
 b_{n-r} &= \beta_1 \sum_{l=2}^{n-r+2} \alpha_l \sum_{s=l+1}^{n-r+3} \alpha_s \dots \sum_{m=u+1}^{n-3} \alpha_m \sum_{k=m+1}^{n-2} \alpha_k \sum_{i=k+1}^{n-1} \alpha_i \sum_{j=i+1}^n \alpha_j - \beta_2 \alpha_1^{(2)} \sum_{s=3}^{n-r+3} \alpha_s \dots \\
 &\quad \sum_{m=u+1}^{n-3} \alpha_m \dots \sum_{j=i-1}^n \alpha_j + \beta_3 \alpha_1^{(2)} \alpha_2^{(3)} \sum_{p=4}^{n-r+4} \alpha_p \dots \sum_{m=n+1}^{n-3} \alpha_m \sum_{k=m+1}^{n-2} a_k \\
 &\quad \sum_{i=k+1}^{n-1} \alpha_i \sum_{j=i+1}^n \alpha_j - \dots - \beta_{r-1} \alpha_1^{(0)} \alpha_2^{(3)} \dots \alpha_{r-2}^{(r-1)} \sum_{i=r+1}^n \alpha_i + \beta_r \alpha_1^{(2)} \alpha_2^{(3)} \dots \alpha_{r-1}^{(r)}
 \end{aligned}$$

El diagrama del filtro de formación, construido en correspondencia con (5.35) se muestra en la Tabla 5.5. Considerando  $B^T = \{0, 0, \dots, \beta_n\}$ ,  $\alpha_i = -\alpha_i^{(i+1)}$ , el filtro obtenido es la conexión en cascada de  $n$  circuitos  $RC$  con constante de tiempo  $1/\alpha_i = R_i C_i$ . Cuando la cantidad de circuitos,  $n$ , es suficientemente grande el filtro se aproxima al tipo Gaussiano.

Modelo del proceso	Función de correlación	Filtro formante
$\frac{dx}{dt} = -\alpha x + \alpha \xi(t),$ $x(t_0) = x_0$	$R_x(\tau) = \sigma_x^2 e^{-\alpha \tau }$	 <p style="text-align: center;"><math>\alpha = 1/RC</math></p>
$\frac{dx_1}{dt} = -\alpha x_1(t) - \alpha x_2(t) + \xi(t),$ $\frac{dx_2}{dt} = -\alpha x_2(t) + \xi(t)$ $x_i(t_0) = x_{i0}, i = 1, 2$	$R_{x_1}(\tau) = \frac{N_x}{8\alpha} (1 - \alpha \tau ) e^{-\alpha \tau }$	 <p style="text-align: center;"><math>\alpha = 1/RC</math></p>

**Figura 5.5.** Modelos de transformación a procesos de Markov

Aunque se tenga un proceso aleatorio real escalar, su representación en forma de un proceso Gaussiano de Markov se realiza en la forma vectorial (5.28), la cual se puede generalizar al caso de análisis conjunto de vectores funcionales continuos con dimensión  $r$ . Cada componente vectorial, a su vez, se modela por el respectivo vector de procesos Gaussianos de Markov  $\mathbf{x}_i(t)$ , descrito por (5.28) en la forma,

$$\frac{d\mathbf{x}_i}{dt} = \mathbf{A}_i(t) \mathbf{x}_i(t) + \mathbf{B}_i(t) \xi_i(t), \quad \mathbf{x}_i(t_0) = \mathbf{x}_{i0}$$

donde  $\mathbf{A}_i$  y  $\mathbf{B}_i$  son matrices, cada una con dimensión  $n_i \times n_i$ , definidas en (5.29). Como resultado, la señal vectorial continua se modela por un proceso Gaussiano  $\mathbf{x}(t)$ , descrito por la ecuación diferencial vectorial y estocástica,

$$\frac{d\mathbf{x}}{dt} = \mathbf{A}(t)\mathbf{x}(t) + \mathbf{B}(t)\boldsymbol{\xi}(t), \quad \mathbf{x}(t_0) = \mathbf{x}_0$$

donde

$$\mathbf{A} = \begin{bmatrix} \mathbf{A}_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{A}_2 & \cdots & \mathbf{0} \\ \cdots & \cdots & \cdots & \cdots \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{A}_r \end{bmatrix}, \quad \mathbf{B} = \begin{bmatrix} \mathbf{B}_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{B}_2 & \cdots & \mathbf{0} \\ \cdots & \cdots & \cdots & \cdots \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{B}_r \end{bmatrix}, \quad \mathbf{x}(t) = \begin{bmatrix} x_1(t) \\ x_2(t) \\ \vdots \\ x_r(t) \end{bmatrix}, \quad \boldsymbol{\xi}(t) = \begin{bmatrix} \xi_1(t) \\ \xi_2(t) \\ \vdots \\ \xi_r(t) \end{bmatrix}$$

La excitación del sistema corresponde a proceso vectorial del tipo RBG  $\boldsymbol{\xi}(t)$  con valor medio  $E\{\boldsymbol{\xi}(t)\} = \mathbf{0}$  y función de correlación  $E\{\boldsymbol{\xi}(t)\boldsymbol{\xi}^\top(t-\tau)\} = \mathbf{R}_\xi(\tau)$ , que es una matriz de varianzas, simétrica y positiva definida, para los ruidos blancos componentes.

#### 5.3.4. Discretización de procesos continuos de Markov

El modelado de procesos continuos dinámicos de Markov en procesadores digitales se realiza mediante la discretización, sobre mallas de puntos determinados en el tiempo, a partir de las ecuaciones diferenciales que describen su comportamiento.

Sea un sistema dinámico, cuyo modelo se describe por la ecuación diferencial,

$$\frac{d\xi}{dt} = f(t, \xi) + g(t, \xi)d\nu(t), \quad \xi(t_0) = \xi_0, \quad t \in T$$

donde  $f$  y  $g$  son funciones determinísticas, continuas y diferenciables, que cumplen la condición de Lipschitz. Al considerar RBG,  $\eta(t)$ , a la entrada de un sistema, entonces su salida es un proceso escalar de Markov, determinado por la ecuación diferencial de Ito:

$$\frac{d\xi}{dt} = f(t, \xi) + g(t, \xi)d\nu(t), \quad \xi(t_0) = \xi_0, \quad t \in T \quad (5.36)$$

en el cual, las funciones  $g$  y  $f$  se consideran monótonas crecientes pero acotadas,

$$|f(t, \xi)|^2 + |g(t, \xi)|^2 \leq c^2(1 + |\xi|^2), \quad 0 < c < \infty, \quad t \in T$$

Las restricciones impuestas a los coeficientes (5.36) son suficientes para la existencia y unicidad de la solución  $\xi(t)$  en el intervalo  $t \in [0, T]$ , con lo cual a la ecuación estocástica diferencial (5.36) le corresponde la representación integral equivalente,

$$\xi(t) = \xi(t_0) + \int_{t_0}^t f(s, \xi(s))ds + \int_{t_0}^t g(s, \xi(s))dsd\nu(s) \quad (5.37)$$

La discretización del modelo del sistema se hace a partir de (5.36) o (5.37), para lo

cual, sobre un intervalo de análisis  $[0, T]$  se genera la malla de puntos uniformes de tiempo  $t_k = k\Delta$ , donde  $\Delta = T/n$ , es el paso de discretización. Luego, dada la condición inicial  $\xi_0 = \xi(0)$ , mediante el principio de estimación recurrente  $\tilde{\xi}_{v+1} = \tilde{\xi}_v + \theta_v(\tilde{\xi}_v, \Delta, \Delta v_v)$  para cada uno de los incrementos del proceso de Wiener,  $\Delta v_v = v(t_{v+1}) - v(t_v)$ , es necesario obtener los valores estimados  $\tilde{\xi}_v$ , que sean los más cercanos, para algún criterio dado a la solución verdadera  $\xi_v = \xi(t_v)$ . Esta cercanía de los valores de  $\xi_v$  y  $\tilde{\xi}_v$ , se da por el valor cuadrático medio del error  $\varepsilon$ , en la forma

$$\varepsilon = \max_{1 \leq v \leq n} \left\{ E\{(\xi_v - \xi_0)^2 | \xi_0\} \right\}^{1/2} \quad (5.38)$$

en el cual, las condiciones iniciales de las soluciones verdadera y aproximada coinciden, esto es,  $\xi_0 = \tilde{\xi}_{01}$ . Asumiendo la suavidad de las funciones  $f$  y  $g$  durante el intervalo de discretización, la expresión (5.37) corresponde a la ecuación iterativa de Euler,

$$\tilde{\xi}_{v+1} = \tilde{\xi} + f(v\Delta t, \tilde{\xi}_v)\Delta + g(v\Delta t, \tilde{\xi}_v)\Delta v \quad (5.39)$$

que en forma general, presenta un error del tipo  $\varepsilon = o(\sqrt{\Delta})$  [42].

La ecuación diferencial estocástica equivalente de Stratanovich para (5.39) tiene la siguiente forma:

$$d\xi = (f(t, \xi) - (1/2)g'_\xi(t, \xi)g(t, \xi))dt + g(t, \xi)dv, \quad \xi(0) = \xi_0 \quad (5.40)$$

donde  $g'_\xi(t, \xi) = \partial g(t, \xi)/\partial \xi$ . La equivalencia de (5.36) y (5.38) se entiende en el sentido, en que ambas expresiones, con iguales condiciones iniciales, conllevan a iguales resultados.

La representación integral equivalente de (5.40) tiene la forma.

$$\xi(t) = \xi(t_0) + \int_{t_0}^t (f(s, \xi(s)) + \frac{1}{2}g'_\xi(s, \xi(s))ds + \int_{t_0}^t g(s, \xi(s))dv(s) \quad (5.41)$$

La ecuación recursiva para (5.41) se expresa como

$$\tilde{\xi}_{v+1} = \tilde{\xi}_v + (f(v\Delta + \tilde{\xi}_v) - (1/2)g'_\xi(v\Delta, \tilde{\xi}_v))\Delta + g(v\Delta + \Delta/2, \tilde{\xi}_v + \Delta\tilde{\xi}_v/2)\Delta v + o(\Delta) \quad (5.42)$$

La aproximación de (5.42) se puede obtener si la función  $g(v\Delta + \Delta/2, \tilde{\xi}_v + \Delta\tilde{\xi}_v/2)$  se descompone en series de potencia sobre el punto  $(v\Delta, \tilde{\xi}_v)$  y se desprecian los términos, cuyo orden sea mayor a  $o(\Delta)$ . Luego (5.42) toma la forma

$$\begin{aligned} \tilde{\xi}_{v+1} = \tilde{\xi}_v + (f(v\Delta, \tilde{\xi}_v) - (1/2)g'_\xi(v\Delta, \tilde{\xi}_v)g(v\Delta, \tilde{\xi}_v))\Delta + g(v\Delta, \tilde{\xi}_v)\Delta v \\ + g'_\xi(v\Delta, \tilde{\xi}_v)(\Delta\tilde{\xi}_v/2)\Delta v + o(\Delta) \end{aligned}$$

Reemplazando en la anterior expresión el incremento,  $\Delta\tilde{\xi} = \tilde{\xi}_{v+1} - \tilde{\xi}_v$ , se obtiene que

$$\tilde{\xi}_{v+1} = \tilde{\xi}_v + f(v\Delta, \tilde{\xi}_v)\Delta + g(v\Delta, \tilde{\xi}_v)\Delta v + (1/2)g'_\xi(v\Delta, \tilde{\xi}_v)(\Delta v^2 + \Delta) + o(\Delta) \quad (5.43)$$



La expresión (5.43) se entiende como una aproximación estocástica mediante la serie de Taylor en cada intervalo de discretización, que en conjunto con la ecuación recursiva (5.41) permiten obtener la solución de la ecuación diferencial estocástica (5.36) de forma recurrente. Así por ejemplo, de (5.42) se obtiene una estructura más simple de cálculo,

$$\xi_{v+1} = \xi_v + f(t_v, \xi_v)\Delta + g(t_v, \xi_v)\Delta v \quad (5.44)$$

Sin embargo, la estructura recursiva obtenida de (5.43)

$$\xi_{v+1} = \xi_v + f(t_v, \xi_v)\Delta + g(t_v, \xi_v)\Delta v + (1/2)g'_\xi(t_v, \xi_v)g(t_v, \xi_v) + (\Delta v^2 - \Delta)$$

presenta menor de error de aproximación, si en (5.44) se asume  $g(t, \xi) = (N/2)^{1/2}$ :

$$\xi_{v+1} = \xi_v + f(t_v, \xi_v) + \sigma_v \eta_v \quad (5.45)$$

donde  $\eta_v$  es RBG con  $E\{\eta_i \eta_j\} = \delta_{ij}$ ,  $\sigma_v^2 = Ng^2(t_v, \xi_v)/2$ . La expresión (5.45) es un algoritmo recurrente de modelado de sucesiones de Markov  $\{\xi_v\}$ , con condición inicial  $\xi_0$ .

**Ejemplo 5.14.** Sea el proceso Gaussiano simple de Markov

$$\frac{d\xi}{dt} = -\alpha\xi + \gamma\eta(t) \quad (1)$$

La correspondiente ecuación recurrente, a partir de (5.45), toma la forma

$$\xi_{v+1} = \xi_v - \alpha\xi_v\Delta + \sigma_v\eta_v, \quad \sigma_v^2 = N\gamma^2\Delta/2 \quad (2)$$

Sin embargo otra forma de discretización de (1) se puede obtener de la solución general

$$\xi(t) = \xi(t_0)\exp(-\alpha(t-t_0)) + \gamma \int_{t_0}^t \exp(-\alpha(t-\tau))\eta(\tau)d\tau$$

con lo cual se obtiene

$$\xi(t) = \exp(-\alpha\Delta)\xi_v + \sigma(\Delta)\eta_v \quad (3)$$

donde

$$\sigma^2(\Delta) = \gamma^2 \frac{N}{2} \int_0^\Delta \exp(-2\tau)d\tau = \frac{\gamma^2 N}{4\alpha} (1 - \exp(-2\alpha\Delta)) \quad (4)$$

La expresión (4) implica que la varianza cumple la relación,  $\alpha\delta^2/d\Delta = -2\alpha\sigma^2 + \gamma^2 N/2$ , con condición inicial  $\sigma^2(0) = 0$ . Las expresiones (3) y (4), a diferencia de (2), establecen la relación precisa entre  $\xi_{v+1}$  y  $\xi_v$ , para cualquiera que sea el valor de  $\Delta$ . Si el proceso de discretización se escoge de tal manera que  $\alpha\Delta \ll 1$ , entonces se pueden emplear la aproximaciones  $\exp(-\alpha\Delta) \approx 1 - \alpha\Delta$  y  $\exp(-2\alpha\Delta) \approx 2\alpha\Delta$ . Luego, ambas expresiones (2) y (3) coinciden, y por lo tanto, la relación (2) será más precisa, entre menor sea el paso  $\Delta$ .

Los resultados anteriores se pueden generalizar para el caso de la ecuación diferencial lineal dada en forma vectorial con parámetros constantes,

$$\dot{\boldsymbol{\xi}} = \mathbf{A}\boldsymbol{\xi}(t) + \boldsymbol{\eta}(t) \quad (5.46)$$

donde  $\boldsymbol{\xi}(t)$  es el proceso vectorial,  $\mathbf{A}$  es una matriz cuadrada y  $\boldsymbol{\eta}(t)$  es el vector RBG con valor medio 0 y matriz de correlación,  $\mathbf{Q}$ , simétrica y positiva semidefinida:

$$E \left\{ \boldsymbol{\eta}(t_1) \boldsymbol{\eta}^\top(t_2) \right\} = \mathbf{Q} \delta(t_2 - t_1) \quad (5.47)$$

La solución general de la ecuación (5.46) tiene la forma

$$\boldsymbol{\xi}(t) = \boldsymbol{\Phi}(t - t_0) \boldsymbol{\xi}(t_0) + \int_{t_0}^t \boldsymbol{\eta}(\boldsymbol{\Phi}(t - \tau) \boldsymbol{\eta}_0(\tau) d\tau \quad (5.48)$$

siendo  $\boldsymbol{\Phi}(t)$  la matriz de transición, que cumple la ecuación:  $\dot{\boldsymbol{\Phi}} = \mathbf{A}\boldsymbol{\Phi}(t)$ , con condición inicial  $\boldsymbol{\Phi}(0) = \mathbf{I}$ . De la ecuación (5.48) y teniendo en cuenta que,

$$\boldsymbol{\Phi}(t) = \exp(\mathbf{A}t) = \mathbf{I} + \mathbf{A}t + (1/2)\mathbf{A}^2 t^2 + \dots \quad (5.49)$$

se obtiene la expresión recursiva  $\boldsymbol{\xi}_{v+1} = \exp(\mathbf{A}\Delta)\boldsymbol{\xi}_v + \boldsymbol{\eta}_v$ . La matriz de correlación del vector ruido blanco Gaussiano se determina, a partir de la expresiones (5.47) y (5.48),

$$\begin{aligned} \sigma^2(\Delta) = E\{\boldsymbol{\eta}_v \boldsymbol{\eta}_v^\top\} &= E \left\{ \left( \int_{v\Delta}^{(v+1)\Delta} \exp(\mathbf{A}((v+1)\Delta - \tau_1)) \boldsymbol{\eta}(\tau_1) d\tau_1 \right) \times \right. \\ &\quad \left. \times \left( \int_{v\Delta}^{(v+1)\Delta} \exp(\mathbf{A}((v+1)\Delta - \tau_2)) \boldsymbol{\eta}(\tau_2) d\tau_2 \right)^\top \right\} \end{aligned}$$

que al introducir el operador de la esperanza se tiene que

$$\begin{aligned} \sigma^2(\Delta) &= \int_0^\Delta \int_0^\Delta \exp(\mathbf{A}^\top \tau_1) E \left\{ \boldsymbol{\eta}((v+1)\Delta - \tau_1) \boldsymbol{\eta}^\top((v+1)\Delta - \tau_2) \right\} \times \\ &\quad \times \exp(\mathbf{A}^\top \tau_2) d\tau_1 d\tau_2 = \int_0^\Delta \exp(\mathbf{A}\tau) \mathbf{Q} \exp(\mathbf{A}^\top \tau) d\tau \end{aligned} \quad (5.50)$$

luego, al multiplicar por la matriz  $\mathbf{A}$  resulta en

$$\mathbf{A}\sigma^2(\Delta) = \mathbf{A} \int_0^\Delta \exp(\mathbf{A}\tau) \mathbf{Q} \exp(\mathbf{A}^\top \tau) d\tau = \int_0^\Delta (\mathbf{A} \exp(\mathbf{A}\tau) d\tau) \mathbf{Q} \exp(\mathbf{A}^\top \tau)$$

En este caso, se puede emplear la integración por partes en forma matricial

$$\begin{aligned}
 \mathbf{A}\boldsymbol{\sigma}^2(\Delta) &= \int_0^\Delta d(\exp(\mathbf{A}\tau)) \mathbf{Q} \exp(\mathbf{A}^\top \tau) d\tau \\
 &= \exp(\mathbf{A}\tau) \mathbf{Q} \exp(\mathbf{A}^\top \tau) \Big|_0^\Delta - \int_0^\Delta \exp(\mathbf{A}\tau) d(\mathbf{Q} \exp(\mathbf{A}^\top \tau)) \\
 &= \exp(\mathbf{A}\Delta) \mathbf{Q} \exp(\mathbf{A}^\top \Delta) - \mathbf{I} \mathbf{Q} \mathbf{I} - \left( \int_0^\Delta (\mathbf{A}\tau) \mathbf{Q} \exp(\mathbf{A}^\top \tau) d\tau \right) \mathbf{A}^\top \\
 &= \frac{d\boldsymbol{\sigma}^2(\Delta)}{d\Delta} - \mathbf{Q} - \boldsymbol{\sigma}^2(\Delta) \mathbf{A}^\top
 \end{aligned}$$

Con lo cual, la matriz  $\boldsymbol{\sigma}^2(\Delta)$  cumple la ecuación

$$\frac{d\boldsymbol{\sigma}^2(\Delta)}{d\Delta} = \mathbf{A}\boldsymbol{\sigma}^2 + \boldsymbol{\sigma}^2 \mathbf{A}^\top + \mathbf{Q} \quad (5.51)$$

con condición inicial  $\boldsymbol{\sigma}^2(0) = 0$ . La matriz  $\boldsymbol{\sigma}^2(\Delta)$  es preferible representarla en forma,

$$\boldsymbol{\sigma}^2 = \mathbf{G} \mathbf{G}^\top \quad (5.52)$$

La matriz  $\mathbf{G}$  es triangular inferior y genera una forma cómoda para el modelado,

$$\boldsymbol{\xi}_{v+1} = \exp(\mathbf{A}\Delta) \boldsymbol{\xi}_v + \mathbf{G} \boldsymbol{\eta}_v = \boldsymbol{\Phi}(\Delta) \boldsymbol{\xi}_v + \mathbf{G} \boldsymbol{\eta}_v \quad (5.53)$$

donde  $E\{\boldsymbol{\eta}_v \boldsymbol{\eta}_\mu^\top\} = \mathbf{I} \delta_{\mu v}$ .

La expresión (5.53) es la generalización vectorial de la forma escalar (3) y permite el modelado del proceso  $\boldsymbol{\xi}(t)$  en tiempo discreto. El algoritmo (5.53) para discretización de procesos continuos de Markov incluye los siguientes pasos:

1. Dada la matriz  $\mathbf{A}$ , mediante métodos de cálculo numérico se resuelve la ecuación  $\dot{\boldsymbol{\Phi}} = \mathbf{A}\boldsymbol{\Phi}(t)$  y se halla la matriz  $\boldsymbol{\Phi}(\Delta) = \exp(\mathbf{A}\Delta)$ .
2. Conocidas  $\mathbf{A}$  y  $\mathbf{Q}$ , por métodos numéricos se resuelve (5.51) y se halla  $\boldsymbol{\sigma}^2(\Delta)$ .
3. De la matriz  $\boldsymbol{\sigma}^2(\Delta)$  se halla la matriz cuadrada  $\mathbf{G}$ .
4. A partir de (5.53), dado  $\boldsymbol{\xi}_v$  se estima el siguiente valor discreto de la sucesión  $\boldsymbol{\xi}_{v+1}$ , para lo cual se emplea RBG, generado por algún método convencional, en calidad de señal de entrada.

En algunos casos, por ejemplo, cuando se tienen procesos vectoriales  $\boldsymbol{\xi}(t)$  en dimensión baja o cuando la matriz  $\mathbf{A}$  es diagonal, es preferible hallar de forma analítica las matrices  $\boldsymbol{\Phi}(t)$  y  $\boldsymbol{\sigma}^2(t)$ . En particular, empleando la descomposición matricial por Taylor de la matriz  $\boldsymbol{\Phi}(t)$ , y asumiendo la diagonalidad de  $\mathbf{A}$  se obtiene  $\exp(\mathbf{A}t) = \text{trace}[\exp(a_1 t) \cdots \exp(a_n t)]$ .

Sin embargo, la descomposición de la misma matriz  $\Phi(t)$  es más cómodo realizarla mediante la transformada de Laplace,  $\widehat{\Phi}(p) = \mathcal{L}\{\Phi(t)\}$ , en la cual cada elemento es la imagen por Laplace del respectivo elemento de la matriz  $\Phi(t)$ . Entonces, de  $\dot{\Phi} = \mathbf{A}\Phi(t)$  se obtiene

$$(p\mathbf{I} - \mathbf{A}) = \widehat{\Phi}(p) = \mathbf{I}, \mathcal{L}\{\Phi(t)\} = p\widehat{\Phi}(p) - \Phi(0) = p\widehat{\Phi}(p) - \mathbf{I}$$

si se tiene que el determinante  $\det(p\mathbf{I} - \mathbf{A}) \neq 0$ , entonces,

$$\Phi(p) = (p\mathbf{I} - \mathbf{A})^{-1} \quad (5.54)$$

que se emplea en el cálculo de la exponente matricial  $\xi(t) = \mathcal{L}^{-1}\{(p\mathbf{I} - \mathbf{A})\}$ . La expresión vectorial para el caso lineal del algoritmo de aproximación escalar, ver ec. (2) del ejemplo 5.14, se obtiene truncando la serie de Taylor para  $t \rightarrow \infty$ , tal que,  $\Phi(t) \simeq \mathbf{I} + \mathbf{A}(t)$ , luego, de (5.50) se obtiene  $\sigma^2(\Delta) \sim \mathbf{A}\Delta$  y la representación vectorial de (5.51) toma la forma

$$\xi_{v+1} = (\mathbf{I} + \mathbf{A}\Delta)\xi_v + \mathbf{\Gamma}_1\eta_v, \quad E\{\eta_\nu\eta_\mu^\top\} = \mathbf{I}\delta_{\mu\nu} \quad (5.55)$$

La matriz diagonal inferior  $\mathbf{\Gamma}_1$  se obtiene de la expresión

$$\mathbf{\Gamma}_1\mathbf{\Gamma}_1^\top = \mathbf{Q}\Delta \quad (5.56)$$

**Ejemplo 5.15.** Basados en el algoritmo de representación matricial del algoritmo de la ec. (2) del ejemplo 5.14, hallar el modelo de una señal de voz, dado por el sistema

$$\frac{d\xi_1}{dt} + \alpha\xi_1 = \frac{d\xi_2}{dt}, \quad \frac{d\xi_2}{dt} + \beta\xi_2 = \eta(t) \quad (1)$$

La representación de (1) en la forma generalizada (5.46) es la siguiente

$$\frac{d\xi_1}{dt} = -\alpha\xi_1 + \beta\xi_2 + \eta(2), \quad \frac{d\xi_2}{dt} = -\beta\xi_2 + \eta(t) \quad (2)$$

Al comparar (5.46) con (2) se obtienen las matrices

$$\mathbf{A} = \begin{bmatrix} -\alpha & -\beta \\ 0 & -\beta \end{bmatrix}, \quad \mathbf{Q} = \frac{N}{2} \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \quad (3)$$

Mediante (5.54), se halla la transformada de Laplace de la matriz  $\Phi(t)$ ,

$$\widehat{\Phi}(p) = (p\mathbf{I} - \mathbf{A})^{-1} = \begin{bmatrix} p + \alpha & \beta \\ 0 & \beta \end{bmatrix}^{-1} = \frac{1}{(p + \alpha)(p + \beta)} \begin{bmatrix} p + \beta & -\beta \\ 0 & p + \beta + \alpha \end{bmatrix} \quad (4)$$

Luego, la matriz de transición obtenida es

$$\Phi(t) = \mathcal{L}^{-1}\{\widehat{\Phi}(p)\} = \begin{bmatrix} \exp(-\alpha t) & \frac{\beta}{\alpha - \beta}(\exp(-\alpha t) - \exp(-\beta t)) \\ 0 & \exp(-\beta t) \end{bmatrix}$$

Basados en los valores (3) y (4), además teniendo en cuenta (5.50), se determinan los valores

de la matriz de varianza  $\sigma^2(\Delta)$  (siendo  $T(x) = 1 - \exp(-\Delta x)$ ):

$$\begin{aligned} d_{11} &= \frac{N_0}{2(\alpha - \beta)^2} \left( \frac{\alpha}{2} T(2\alpha) - \frac{\alpha\beta}{\alpha + \beta} T(\alpha + \beta) + \frac{\beta}{2} T(2\beta) \right) \\ d_{12} &= \frac{N_0}{2(\alpha - \beta)^2} \left( \frac{\alpha}{2} T(\alpha + \beta) - \frac{1}{1(\alpha + \beta)} T(2\beta) \right) \\ d_{22} &= \frac{1}{2\beta} T(2\beta) \end{aligned} \quad (5)$$

Por lo anterior, un proceso vectorial aleatorio de Markov  $\xi^T = \{\xi_1, \xi_2\}$  se modela mediante el algoritmo (5.53), si se descompone la matriz  $\sigma^2(\Delta)$  con elementos (5) de acuerdo con la expresión (5.52). El algoritmo de aproximación (5.55) se puede obtener de dos formas: empleando la expresión (5.56) o hallando la primera aproximación, cuando  $\Delta \rightarrow 0$ , para la expresión (5). En ambos casos se obtiene el mismo resultado.

Luego

$$\Phi(\Delta) \sim \begin{bmatrix} 1 - \alpha\Delta & 1 - \beta\Delta \\ 0 & 1 - \beta\Delta \end{bmatrix}, \quad \sigma^2(\Delta) \sim \frac{N\Delta}{2} \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$$

El reemplazo de los valores respectivos implica que la solución de la ecuación  $\Gamma_1 \Gamma_1^T = \sigma^2(\Delta)$  corresponde a la matriz

$$\Gamma_1 = \left( \frac{N\Delta}{2} \right)^{1/2} \begin{bmatrix} 1 & 0 \\ 1 & 0 \end{bmatrix}$$

El algoritmo de aproximación (5.53) para el ejemplo en análisis tiene la forma

$$\xi_{v+1} = \begin{bmatrix} 1 - \alpha t & \beta \Delta t \\ 0 & 1 - \beta \Delta t \end{bmatrix} \xi_v + \left( \frac{N\Delta}{2} \right)^{1/2} \begin{bmatrix} 1 & 0 \\ 1 & 0 \end{bmatrix} \eta_v$$

**Ejemplo 5.16.** Sea un proceso bicomponente Gaussiano de Markov en la forma

$$\frac{\xi_1}{dt} = -\alpha_1 \xi_1 - \beta_1 \xi_2 + \eta_1(t), \quad \frac{\xi_2}{dt} = -\alpha_2 \xi_1 - \beta_2 \xi_2 + \eta_2(t)$$

$$\text{donde } E\{\eta_i(t)\} = 0, \quad E\{\eta(t_1)\eta^T(t_2)\} = \begin{bmatrix} \frac{N}{2} & r \frac{\sqrt{N_1 N_2}}{2} \\ r \frac{\sqrt{N_1 N_2}}{2} & \frac{N_2}{2} \end{bmatrix}.$$

La matriz  $\mathbf{A}$ , en este caso, de determina como

$$\mathbf{A} = \begin{bmatrix} -\alpha_1 & -\beta_1 \\ -\alpha_2 & -\beta_2 \end{bmatrix}$$

El cálculo de la matriz  $\Phi$  se puede realizar mediante el método de diagonalización. En estos casos, inicialmente se calculan los valores propios  $\lambda_{1,2}$  de la matriz  $\mathbf{A}$ , como la solución de la ecuación  $\det(\lambda \mathbf{I} - \mathbf{A}) = 0$ :

$$\lambda_{1,2} = -(\alpha_1 + \beta_2 \mp \omega)/2, \quad \omega = ((\alpha_1 - \beta_2)^2 + 4\alpha_2\beta_1)^{1/2}$$

asumiendo que  $\alpha_{1,2} \in \mathbb{R}$ , entonces existe una matriz regular  $\mathbf{C}$  que transforme a  $\mathbf{A}$  a la forma

diagonal, esto es,

$$\mathbf{C}\mathbf{A}\mathbf{C}^{-1} = \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix} = \mathbf{R} \quad (6)$$

Los elementos de la matriz  $\mathbf{C}$  se hallan de la ecuación obtenida de la multiplicación de la igualdad (6) por  $\mathbf{C}$ :

$$\mathbf{A} = \mathbf{R}\mathbf{C}, \Rightarrow \mathbf{C} = \begin{bmatrix} \nu/2\beta_1 & 1 \\ 1 & -\nu/2\alpha_2 \end{bmatrix}, \mathbf{C}^{-1} = 2 \begin{bmatrix} \nu\beta_1/\delta & 2\alpha_2\beta_1/\delta \\ 2\alpha_2\beta_1/\delta & -\nu\alpha_2/\delta \end{bmatrix}$$

donde  $\nu = \alpha_1 - \beta_2 - \omega$ ,  $\delta = \nu^2 + \alpha_2\beta_1$ .

De (6), se tiene que  $\mathbf{A} = \mathbf{C}^{-1}\mathbf{R}\mathbf{C}$ , por lo tanto,

$$\mathbf{A}^n = (\mathbf{C}^{-1}\mathbf{R})(\mathbf{C}^{-1}\mathbf{R}\mathbf{A}) \dots (\mathbf{A}^{-1}\mathbf{R}\mathbf{C}) = \mathbf{C}^{-1}\mathbf{R}^n\mathbf{C}$$

El empleo de la descomposición (5.49) de la matriz

$$\Phi(t) = \exp(\mathbf{A}t) = \mathbf{I} + \mathbf{A}t + (1/2)\mathbf{A}^2t^2 + \dots + \mathbf{C}^{-1}\mathbf{I}\mathbf{C} + \mathbf{C}^{-1}\mathbf{R}t\mathbf{C} + \dots \quad (7)$$

$$= \mathbf{C}^{-1}(\mathbf{I} + \mathbf{R}t + (1/2)\mathbf{R}^2t^2 + \dots)\mathbf{C} = \mathbf{C}^{-1}\exp(\mathbf{R}t)\mathbf{C} \quad (5.57)$$

Por cuanto la matriz  $\mathbf{R}$  es ortogonal, entonces  $\exp(\mathbf{A}t) = \begin{bmatrix} \exp(\lambda_1 t) & 0 \\ 0 & \exp(\lambda_2 t) \end{bmatrix}$

Teniendo en cuenta las expresiones de  $\mathbf{C}$  y  $\mathbf{C}^{-1}$ , además la relación (7), se halla que

$$\phi(\Delta) = \begin{bmatrix} \nu^2 \exp(\lambda_1 \Delta) + 4\lambda_2 \beta_1 \exp(\lambda_2 \Delta) & 2\nu\lambda_2 (\exp(\lambda_1 \Delta) - \exp(\lambda_2 \Delta)) \\ 2\nu\beta_1 (\exp(\lambda_1 \Delta) - \exp(\lambda_2 \Delta)) & 4\lambda_2 \beta_1 \exp(\lambda_1 \Delta) + \nu^2 \exp(\lambda_2 \Delta) \end{bmatrix}$$

Los elementos de la matriz de varianza  $\sigma^2(\Delta)$  se muestran en [41].

## Problemas

**Problema 5.17.** La ecuación diferencial de un generador de señales senoidales tiene la forma

$$\dot{a} - ba \left(1 - \frac{a^2}{4}\right) = c\xi(t)$$

donde  $a$  es la amplitud,  $b, c = \text{const}$  y  $\xi(t)$  son las fluctuaciones que generan cambios en la amplitud. Hallar la ecuación diferencial lineal con relación al incremento relativo de amplitud y su solución estacionaria, empleando el método de linealización determinística. Asumir  $E\{\xi(t)\} = 0$  y  $\sigma_a \ll a_0$ .

**Problema 5.18.** Hallar el respectivo modelo de la señal de un proceso  $\xi(t)$ , a la salida de un circuito de resonancia, dado por la siguiente ecuación lineal diferencial de segundo orden:

$$d^2\xi/dt^2 + 2\alpha d\xi/dt + \omega_0^2\xi = \omega_0^2\eta(t), \quad \omega_0 \gg \alpha$$

# Capítulo 6

## Detección y filtración de señales aleatorias

Dada una mezcla de la señal útil  $\xi$  y alguna perturbación  $\eta$  presente en ella, el objetivo de la filtración consiste en obtener la estimación, bien sea de la misma señal útil en general,  $\tilde{\xi}$ , o bien de alguno de sus parámetros o características de interés,  $\xi(\tilde{\theta})$ , con estructura aleatoria. De otra manera, la filtración implica buscar la transformación de la trayectoria  $\mathbf{x}(t) \in \xi$  en análisis a fin de obtener la mejor estimación, por un criterio dado, de los valores de un proceso aleatorio en un momento determinado del tiempo del intervalo de observación. Además de la estimación del mismo proceso  $\xi$ , en la práctica, a partir de las respectivas observaciones  $\mathbf{x} \in \xi$ , dadas sobre un intervalo de tiempo de análisis, tiene sentido la estimación de alguna transformación lineal del proceso aleatorio, por ejemplo, el desplazamiento en el tiempo, la diferenciación o integración múltiple o la combinación de estas transformaciones.

Aunque en la mayoría de los casos, los sistemas óptimos de proceso son no lineales. No obstante, el desarrollo de los métodos de filtración lineal tiene mayor importancia, por cuanto su aplicación presenta mejor sentido práctico. De otra parte, debido a que la condición de linealidad no implica la condición de realización física (causalidad del sistema), es usual para la síntesis de filtros el empleo de la solución secuencial: primero, la comprobación de la condición de linealidad y luego, la de causalidad.

Como criterio de calidad en la estimación del valor de la señal aleatoria, en algunos casos, se escoge el valor medio de la potencia del error para un ensamble dado de trayectorias, esto es, el valor cuadrático medio de la desviación estándar de la estimación del valor estimado. En otros casos, el criterio de calidad corresponde a la relación de las potencias de la señal contra la de perturbación.

El análisis de filtros no lineales se puede considerar de diversas maneras; cada una de ellas acopladas a la naturaleza de no linealidad del proceso. En general, entre las aproximaciones más conocidas, descritas en el numeral §5.2.1, están las relacionadas con la caracterización de los sistemas no lineales mediante la suma infinita de los integrales de Volterra. Otra aproximación consiste en la representación, dadas las restricciones pertinentes, en forma de procesos de Markov.

## 6.1. Métodos de detección

Sea el conjunto de señales de información,  $\{\mathbf{x}_k(t) : k = 1, \dots, n\}$  con estructura conocida, bien sea de naturaleza determinística o aleatoria, y de las cuales solo puede existir una señal,  $\mathbf{x}_k(t)$ , durante un intervalo dado de observación  $T$ , que además se asume perturbada por alguna trayectoria de la distorsión aleatoria  $\eta(t)$ .

A efectos de análisis, la perturbación  $\eta(t)$  se considera RBG aditivo, con lo cual, la trayectoria sobre la cual se realiza la detección tiene la forma:

$$\mathbf{y}(t) = \mathbf{x}_k(t) + \eta(t), \quad t \in T \quad (6.1)$$

De tal manera, que pueden generarse las respectivas hipótesis sobre cuál de las  $k$  posibles señales está presente en el intervalo dado de observación.

La tarea de detección, de acuerdo con el número  $k$  de posibles señales de información, puede ser binaria, ( $k = 2$ ), o bien múltiple, ( $k > 2$ ). Sobre el mismo carácter de las señales de información, se pueden diferenciar las siguientes tres tareas:

- (a). Detección de las señales determinísticas con parámetros conocidos (sistemas sincrónicos de comunicación digital).
- (b). Detección de señales determinísticas con parámetros desconocidos (detección de radar, sistemas asincrónicos de comunicación digital).
- (c). Detección de señales de información con estructura aleatoria (detección sísmica, radioastronomía, radioprospección).

Sea un conjunto de señales  $\mathbf{x}_k(t)$ ,  $k = 1, \dots, N$ , dado el modelo de observación (6.1), para el cual se introduce la FDP condicional con dimensión múltiple,  $p(\mathbf{y}|\mathbf{x}_k)$  de la variable  $\mathbf{y}$ , bajo la condición de que esté presente la señal de información  $\mathbf{x}_k$ , en el intervalo dado de observación. La toma de decisión sobre cuál de las posibles señales está presente, implica que todo el espacio  $\mathfrak{G}$  conformado por las señales de información sea dividido de forma determinada en  $m$  subespacios,  $\{G_i \subset \mathfrak{G} : i = 1, \dots, N\}$ . De tal manera, que al tenerse  $\mathbf{y}(t) \in G_i$ ,  $t \in T$ , entonces, se toma la decisión de que está presente la señal  $\mathbf{x}_i(t)$ .

Sin embargo, debido a la interacción del ruido  $\eta(t)$ , el verdadero valor de la medida de la señal presente  $\mathbf{x}_k(t)$  se altera, como se muestra en la Figura 6.1, luego, pueden ocurrir dos casos extremos en la toma de decisión:

- Primero, que la perturbación  $\eta_1(t)$  cambie el valor real de medida, pero la señal  $\mathbf{x}_k(t)$  siga reflejándose, mediante  $\mathbf{y}_1(t)$ , en su respectivo subespacio  $G_k$ , con lo que la decisión se toma correctamente sobre cuál de las señales fue detectada.
- Segundo, que la trayectoria  $\eta_1(t)$  altere la medida, de tal manera que la señal  $\mathbf{x}_k(t)$ , reflejada en  $\mathbf{y}_2(t)$ , traspase la frontera  $\gamma_{ki}$  y se ubique en un subespacio diferente, esto es,  $\mathbf{x}_k \rightarrow G_i$ , para todo  $i = 1, \dots, N$ ,  $i \neq k$ ; por lo tanto, se genera un error en la toma de decisión durante la detección de la respectiva señal.



Es evidente, que la conformación de los subespacios y, en particular, el cambio de las fronteras influyen sobre la probabilidad de error de la detección de las señales de información. Así por ejemplo, si en vez de tomar la frontera  $\gamma_{km}$ , dada la observación  $y_1$ , entonces la detección de  $x_k(t)$  es incorrecta, por cuanto se toma la decisión  $G_m$ . Por el contrario, para la trayectoria  $y_2$ , al cambiar la frontera hasta  $\gamma_{ki}$  el espacio de decisión  $G_i$  se amplía y la detección de  $x_k$  ya se convierte en correcta.

La FDP condicional de detección correcta se da por la integral múltiple:

$$p(x_i|x_i) = \int_{G_i \in \mathcal{G}} p(y|x_i) dy \quad (6.2)$$

Mientras, la FDP condicional complementaria de error tiene la forma

$$p(x_j|x_i) = \int_{G_i} p(y|x_i) dy, \quad \forall j = 1, \dots, N, \quad j \neq i \quad (6.3)$$

De esta manera, dado un criterio de decisión, la mejor división de los subespacios de las señales de información (toma de decisión óptima de detección) se realiza mediante los métodos de teoría de toma de decisión estadística, descritos en el numeral §2.3.

### 6.1.1. Detección bayesiana de señales

Sea la función de pérdida  $f_c\{\varepsilon\}$  en forma de los valores  $c_{ji}$ , que resultan de la toma incorrecta de la decisión de detección de la señal  $x_i$ , cuando realmente se tenía  $x_j$ . Las funciones de pérdida asumen  $c_{ij} = 0$ , sin embargo, se puede generalizar su definición y considerar el caso cuando  $c_{ij} \neq 0$ , asumiendo el sentido de *función de costos* de decisión.

La función de riesgo condicional, de (2.81), se determina para la detección de  $x_i$  como

$$r_g(x_i) = \sum_{n=1}^N c_{ij} p(x_j|x_i)$$

Dada la FDP a priori de aparición,  $p(x_i)$ , para la señal  $x_i$ , entonces el riesgo medio (2.86), cuando se asume la presencia de una sola señal de información en el intervalo de análisis, se determina como

$$R_g(x_i) = \sum_{n=1}^N r_g(x_i) p(x_i) = \sum_{n=1}^N \sum_{n=1}^N c_{ij} p(x_i) p(x_j|x_i)$$

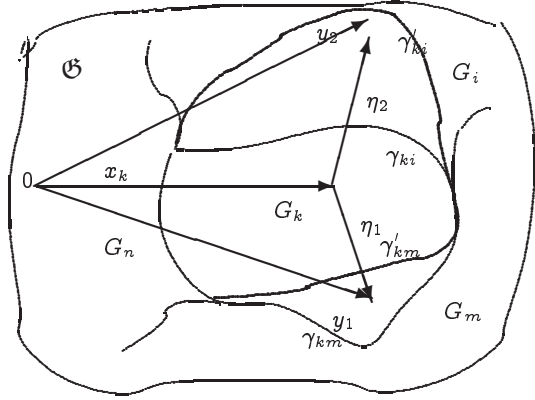


Figura 6.1. Espacios de decisión

Teniendo en cuenta (6.2) y (6.3) se obtiene

$$R_g(x_i) = \sum_{n=1}^N \sum_{n=1}^N c_{ij} p(x_i) \int D_j p(y|x_i) dy \quad (6.4)$$

El método de detección se considera óptimo, de acuerdo al criterio de riesgo medio, cuando el valor de (6.4) se hace mínimo, cuyas variables de optimización son las características de las señales, en particular la estructura y parámetros de los operadores de transformación de las señales mismas, que conforman las fronteras de los respectivos subespacios de toma de decisión.

### Detección binaria

Sea una observación,  $y(t)$ , que corresponde al modelo (6.1), la cual se representa por una sucesión de valores  $y[n]$ ,  $n = 0, \dots, N-1$ , que corresponde a un vector,  $\mathbf{y}$ , con dimensión  $N$ . Sean las hipótesis  $H_0$  y  $H_1$ , conformadas para los casos de aparición de las señales  $\mathbf{x}_0$  y  $\mathbf{x}_1$ , respectivamente, con lo cual se pueden tener los siguientes resultados de decisión:

1. Se acepta  $H_0$  y  $H_0$  es cierta, entonces se asigna el costo de decisión  $c_{00}$
2. Se acepta  $H_1$ , pero  $H_1$  es cierta,  $c_{10}$
3. Se acepta  $H_1$  y  $H_1$  es cierta,  $c_{11}$
4. Se acepta  $H_0$ , pero  $H_1$  es cierta,  $c_{01}$

Las decisiones 1 y 3 son correctas, mientras las decisiones 2 y 4 resultan incorrectas y generan el correspondiente error.

Definidas las probabilidades a priori  $P_0$  y  $P_1$ , que corresponden a la probabilidad de aparición de las señales  $\mathbf{x}_0$  y  $\mathbf{x}_1$ , respectivamente, el riesgo medio (6.4) resulta en

$$R_g = c_{00}p(\mathbf{x}_0)p(\mathbf{x}_0|\mathbf{x}_0) + c_{10}p(\mathbf{x}_0)p(\mathbf{x}_1|\mathbf{x}_0) + c_{11}p(\mathbf{x}_1)p(\mathbf{x}_1|\mathbf{x}_1) + c_{01}p(\mathbf{x}_1)p(\mathbf{x}_0|\mathbf{x}_1) \quad (6.5)$$

El valor de la función de costos se define de tal manera que, los coeficientes de error sean mayores que cero,  $c_{10} > 0, c_{01} > 0$ , mientras para los coeficientes de decisión acertada se tenga,  $c_{00} \leq 0, c_{11} \leq 0$ , con lo cual se cumple que,

$$c_{10} > c_{00}, c_{01} > c_{11}$$

La función de decisión consiste en la conformación, a partir del espacio total de decisión y dada la función de transformación del espacio de observaciones,  $D = g(y)$ , para los subespacios de decisión  $D_0$  y  $D_1$ , que corresponden a las respectivas hipótesis:  $H_0$  y  $H_1$ .

Como antes se dijo, la regla de decisión óptima consiste en la determinación de la frontera  $\gamma$ , que minimice el riesgo medio (6.4). Por cuanto, para el subespacio  $D_i$ ,  $i = 0, 1$ , se toma

la hipótesis  $H_i$ , entonces

$$p(x_0|x_0) = \int_{D_0} p(y|x_0) dy \quad (6.6a)$$

$$p(x_1|x_0) = \int_{D_1} p(y|x_0) dy \quad (6.6b)$$

$$p(x_1|x_1) = \int_{D_1} p(y|x_1) dy \quad (6.6c)$$

$$p(x_0|x_1) = \int_{D_0} p(y|x_1) dy \quad (6.6d)$$

donde  $p(y|x_0)$  y  $p(y|x_1)$  son la FDP condicionales de aparición de la trayectoria  $y(t)$ , cuando se asume la presencia de las señales de información,  $x_0$  y  $x_1$ , respectivamente.

Las integrales de ambas FDP, tomadas por los respectivos subespacios de decisión,  $D_0$  y  $D_1$ , determinan la validez de las correspondientes hipótesis,  $H_0$  y  $H_1$ . Reemplazando en (6.5), las expresiones (6.5) y (6.6d), se obtiene el valor de riesgo

$$\begin{aligned} R = c_{00}p(x_0) \int_{D_0} p(y|x_0) dy + c_{10}p(x_0) \int_{D_1} p(y|x_0) dy \\ + c_{11}p(x_1) \int_{D_1} p(y|x_1) dy + c_{01}p(x_1) \int_{D_0} p(y|x_1) dy \end{aligned}$$

Las relaciones (6.2) y (6.3) son complementarias, en el sentido en que,

$$\int_{D_1} p(y|x_i) dy = 1 - \int_{D_0} p(y|x_i) dy, \quad i = 0, 1 \quad (6.7)$$

luego, el riesgo se expresa en función de un solo subespacio de decisión, por ejemplo,  $D_0$ ,

$$\begin{aligned} R = p(x_0)c_{10} + p(x_0)c_{11} \\ + \int_{D_0} (p(x_1)(c_{01} - c_{11})p(y|x_1) - (p(x_0)(c_{10} - c_{00}))p(y|x_0)) dy \quad (6.8) \end{aligned}$$

Los dos primeros términos en (6.8) corresponden a los valores fijos de pérdida, que no dependen de la selección de las fronteras del subespacio de decisión  $D_0$ . Por lo tanto, la minimización del riesgo  $R$  significa configurar el subespacio  $D_0$ , de tal manera que se incluyan únicamente los valores negativos (valores de pérdida) de la integral en la expresión (6.8), esto es, de la condición:

$$p(x_1)(c_{01} - c_{11})p(y|x_1) - p(x_0)(c_{10} - c_{00})p(y|x_0) < 0$$

que se puede escribir en la forma

$$\frac{p(y|x_1)}{p(y|x_2)} < \frac{p(x_0)(c_{10} - c_{00})}{p(x_1)(c_{01} - c_{11})}$$

De lo anterior, la función de decisión toma la forma

$$\frac{p(y|x_1)}{p(y|x_2)} \underset{H_0}{\overset{H_1}{>}} \frac{p(x_0)(c_{10} - c_{00})}{p(x_1)(c_{01} - c_{11})} = \gamma \quad (6.9)$$

La parte derecha de la expresión (6.9) es la función de verosimilitud,

$$\Lambda(y) \triangleq \frac{p(x_0)(c_{10} - c_{00})}{p(x_1)(c_{01} - c_{11})}$$

que implica en (6.9) que es más verosímil la hipótesis con la mayor FDP a posteriori, para una trayectoria dada de  $y(t)$ .

La parte derecha de (6.9) corresponde a un valor fijo  $\gamma$ , calculado a partir de la relación entre las probabilidades de aparición de las señales  $x_0$  y  $x_1$  y los respectivos costos de decisión, que se determina como

$$\gamma = \frac{p(x_0)(c_{10} - c_{00})}{p(x_1)(c_{01} - c_{11})}$$

En general, el riesgo medio es uno de los criterios de decisión más generalizados, pero su empleo implica una gran cantidad de información sobre los procesos de análisis, que en la práctica no siempre está disponible.

### Detección del observador ideal

Sean los coeficientes de la función de costos iguales y definidos de la forma:

$$c_{ij} = \begin{cases} 0, & i = j \\ 1, & i \neq j \end{cases}$$

con lo cual se asume que todos los errores de decisión conllevan a un mismo costo y, por lo tanto, el riesgo medio corresponde a la probabilidad total de aparición del error

$$R = P_e = \sum_{i=1}^N \sum_{\substack{j=1 \\ j \neq i}}^N p(x_i) \int_{D_j} p(y|x_i) dy \quad (6.10)$$

El criterio (6.10) determina el mínimo valor medio de errores, mientras la probabilidad de detección correcta sea la máxima, que se alcanza cuando la decisión de que la señal detectada pertenece al subespacio  $D_i$ , cumple la condición,

$$p(y|x_i)p(x_i) > p(y|x_j)p(x_j), \quad \forall j \neq i \quad (6.11)$$

En esencia, dadas las  $N - 1$  condiciones (6.11), se tiene la siguiente regla de detección:

$$\max_j \{p(x_j)p(y|x_i)\} = p(x_i)p(y|x_i) \quad (6.12)$$

esto es, la decisión sobre cuál fue la señal detectada recae en aquella que tenga la máxima FDP a priori.

### Detección del máximo de probabilidad a posteriori

Debido a que  $p(x_i)p(y|x_i) = p(x_i|y)p(y)$ , donde  $p(x_i|y)$  es la FDP a posteriori de que se tiene la señal  $x_i$ , cuando se mide la observación  $y$ , mientras,  $p(y)$  es la densidad marginal de la señal medida, entonces, en concordancia con el teorema de Bayes (2.4), se tiene,

$$p(x_i|y) = \frac{p(x_i)p(y|x_i)}{\sum_{i=1}^N p(x_i)p(y|x_i)} \quad (6.13)$$

luego, el algoritmo de detección óptimo, en este caso, tiene la forma

$$\max_j \{p(x_i|y)\} = p(x_i|y) \quad (6.14)$$

La comparación de los algoritmos (6.12) y (6.14) muestra que ambos conllevan a la misma decisión óptima, y por lo tanto, los correspondientes criterios son equivalentes. Además, ambos exigen la misma información a priori del proceso en análisis.

### Detección minimax

En caso de desconocer las probabilidades a priori de aparición de las señales, el riesgo (6.4) se puede calcular sobre la peor densidad de probabilidad,  $p(x_i)$ ,  $i = 1, \dots, N$ , que se asume de tal manera que se obtenga el mayor valor del riesgo medio, luego,

$$R = \min_{D_j} \max_{p(x_i)} \sum_{i=1}^N \sum_{j=1}^N c_{ij} p(x_i) \int_{D_j} p(z|x_i) dz$$

### Detección de Neymann-Pearson

Existen casos en los cuales se tiene una clara asimetría en las pérdidas de decisión. En el caso particular de la detección binaria, el criterio minimiza uno de los dos tipos de error indicados en el ejemplo 2.27.

En la práctica, tiene sentido brindar la mínima probabilidad de error del tipo 1,

$$p(x_0|x_1) = \int_{D_0} p(y|x_1) dy$$

para un valor dado probabilidad de error tipo 2,

$$p(x_1|x_0) = \int_{D_1} p(y|x_0) dy$$

con lo cual, el algoritmo de detección (6.14), para el caso de detección binaria, se simplifica significativamente hasta la expresión  $p(x_1|y) > p(x_0|y)$ , que al tomar en cuenta (6.13) puede describirse de cualquiera de las siguientes dos formas:

$$\begin{aligned} p(x_1)p(y|x_1) &> p(x_0)p(y|x_0) \\ p(y|x_1)/p(y|x_0) &> p(x_0)/p(x_1) \end{aligned}$$

Esto es,  $\Lambda(y) = \gamma$ , donde  $\gamma = p(x_0)/p(x_1)$ . El umbral  $\gamma$  no es conocido, pero se escoge a partir de un valor asumido para el error  $\alpha_2$  del tipo 2. Ambas FDP de error se expresan en función de la relación de verosimilitud, haciendo el cambio de variables,  $y \rightarrow \Lambda$ ,

$$p(y|x_1)dy = p(\Lambda|x_1)d\Lambda, \quad p(y|x_0)dy = p(\Lambda|x_0)d\Lambda$$

En este caso, el espacio de decisión  $D$  se transforma en un eje de valores,  $\Lambda$ , en el cual el valor  $\gamma$  corresponde a la frontera entre los espacios de decisión, por lo tanto,

$$\begin{aligned} p(x_0|x_1) &= \int_{D_0} p(y|x_1)dy = \int_0^{\gamma} p(\Lambda|x_1)d\Lambda \\ p(x_1|x_0) &= \int_{D_1} p(y|x_0)dy = \int_{\gamma}^{\infty} p(\Lambda|x_0)d\Lambda \end{aligned}$$

El valor  $\gamma$ , entonces, se puede determinar dada la condición  $\int_{\gamma}^{\infty} p(\Lambda|x_0)d\Lambda = \alpha_2$ . Como resultado, el detector óptimo por Neymann-Pearson realiza el algoritmo

$$\Lambda > \gamma \tag{6.15}$$

Si la condición (6.15) es cierta, entonces, el algoritmo asume la detección de  $x_1(t)$ .

### Detección por medidas de información

En este caso, el criterio analiza la cantidad de información en la detección, para lo cual se emplea la medida (2.57) que relaciona la información mutua de una señal dada con respecto a los demás, sobre su entropía, (2.58). En el caso binario, este criterio de detección escoge la señal con mayor relación,

$$R = I(x_j, x_i)/H(x_i) = 1 - H(x_i|x_j)/H(x_i), \quad i \neq j, i, j = 0, 1$$

Cabe anotar, que del ejemplo 2.18, dados dos sucesos con probabilidades respectivas  $p_1$

y  $p_2$ , la mayor entropía se obtiene para  $p_1 = p_2$ :

$$H(x_1|x_0) = -(p_1 \ln p_1 + (1 - p_1) \ln (1 - p_1))$$

En la práctica, el criterio de detección informativo conlleva a los mismos resultados que los criterios (6.12) y (6.13).

### 6.1.2. Detección de máxima verosimilitud

Si en el algoritmo (6.12), se asume la FDP uniforme para todas las señales,  $p(x_j) = 1/N$ , entonces el criterio se expresa solamente en términos de la función de verosimilitud,

$$p(y|x_i) = \max_j \{p(y|x_j)\} \quad (6.16)$$

El criterio de detección (6.16) ha obtenido mayor aceptación debido a la simplicidad relativa de implementación, que exige espacios de observación no muy grandes.

De otra parte, teniendo en cuenta (6.9) el criterio de Bayes se puede expresar en términos del criterio de máxima verosimilitud

$$\Lambda(y) \underset{H_0}{\overset{H_1}{\gtrless}} \gamma \quad (6.17)$$

Los valores de las densidades a priori y de los costos influyen solamente en el cálculo del umbral, mas no en la estructura del algoritmo, que se basa en la estimación de  $\Lambda(y)$ . En este sentido la sintonización de los valores de los costos y probabilidades de generación de las señales, que frecuentemente se hace de forma heurística, no influye en la forma de toma de decisión, y más bien ocurre la sintonización fina de la ubicación del umbral.

Debido a que ambas partes de la igualdad (6.17) son positivas, entonces en vez de tomar directamente la relación de verosimilitud, se emplea su logaritmo, que es una función monótona. Así, el criterio (6.17) toma la forma

$$\ln \Lambda(y) \underset{H_0}{\overset{H_1}{\gtrless}} \ln \gamma \quad (6.18)$$

que es adecuada, debido a que en la práctica la mayoría de FDP  $p(y|x_i)$  son exponenciales.

**Ejemplo 6.1.** Desarrollar el criterio de máxima verosimilitud para la detección binaria de las señales  $x_0(t)$  y  $x_1(t)$ , afectadas por el ruido blanco Gaussiano  $\eta(t)$  con parámetros  $\mathcal{N}(0, \sigma_\eta)$ ; asumiendo que el ancho de banda del espectro de las señales es  $\Delta f$ .

Sea la trayectoria del ruido  $\eta(t)$  que se discretiza, con periodo de muestreo  $\Delta t = 1/2\Delta f$ , y conforma la sucesión  $\{\eta_i : i = 1, m\}$ , en el cual se considera que todos sus valores son variables aleatorias independientes. Por lo tanto, la correspondiente FDP de dimensión múltiple Gaussiana se puede representar en la forma,

$$P_\eta(\eta_1, \eta_2, \dots, \eta_m) = \frac{1}{(2\pi\sigma_\eta^2)^{m/2}} \prod_{k=1}^m \exp(-\eta_k^2/2\sigma_\eta^2) = \frac{1}{(2\pi\sigma_\eta^2)^{m/2}} \exp\left(-\frac{1}{2\sigma_\eta^2} \sum_{k=1}^m \eta_k^2\right)$$

donde  $m = T/\Delta t = 2T\Delta f$ , siendo  $T$  el intervalo de análisis para la detección.

La FDP condicional  $p(y|x_i)$  de aparición de la trayectoria de la medida,  $y(t) = x_i(t) + \eta(t)$ , corresponde a la densidad de probabilidad,  $P_\eta(y(t) - x_i(t))$ , de la observación del ruido  $\eta(t) = y(t) - x_i(t)$ , que conforman la sucesión de valores discretizados,

$$\eta_k \neq \eta(t_k) = y(t_k) - x_i(t_k) = \Delta y_i[k]$$

que le corresponde la FDP con dimensión múltiple de la forma

$$p(y|x_i) = P_\epsilon(\Delta y_i[1], \dots, \Delta y_i[m]) = \frac{1}{(2\pi\sigma_\eta^2)^{m/2}} \exp\left(-\frac{1}{2\sigma_\eta^2} \sum_{k=1}^m \Delta y_i^2[k]\right)$$

entonces, la relación de verosimilitud tiene la forma

$$\begin{aligned} \lambda(y) &= \frac{P_\eta(y - x_1)}{P_\eta(y - x_0)} = \exp\left(-\frac{1}{2\sigma_\eta^2} \sum_{k=1}^m (\Delta y_i^2[k] - \Delta_0^2[k])\right) \\ &= \exp\left(-\frac{1}{N_0} \sum_{k=1}^m ((y[k] - x_1[k])^2 - (y[k] - x_0[k])^2) \Delta t\right) \end{aligned} \quad (1)$$

donde  $N_0 = \sigma_\eta^2/\Delta f = 2\sigma_\eta^2\Delta t$  es la densidad espectral de potencia del ruido.

Teniendo en cuenta (1), el criterio de detección (6.18), toma la forma

$$\ln \lambda(y) = -\frac{1}{N_0} \sum_{k=1}^m ((y[k] - x_1[k])^2 - (y[k] - x_0[k])^2) \Delta t \underset{H_0}{\overset{H_1}{\gtrless}} \ln \gamma$$

asumiendo que el ancho de banda  $\Delta f$  es suficientemente grande, y por lo tanto,  $\Delta t \rightarrow 0$ , entonces la sumatoria anterior se reemplaza por la operación de integración,

$$-\frac{1}{N_0} \int_0^T ((y(t) - x_1(t))^2 - (y(t) - x_0(t))^2) dt \underset{H_0}{\overset{H_1}{\gtrless}} \ln \gamma \quad (2)$$

## Problemas

**Problema 6.1.** Desarrollar el algoritmo de detección, dado el RBG con valor medio cero y matriz de correlación  $\mathbf{R}$ , para el proceso aleatorio descrito en la forma:

$$y(t, \vartheta) = \sum_{k=1}^m \vartheta_k x_k(t) = \vartheta \mathbf{x}(t)$$

donde  $\mathbf{x}(t) = \{x_k(t) : k = 1, \dots, m\}$  es un sistema dado de funciones determinísticas linealmente independientes,  $\vartheta = \{\vartheta_k : k = 1, \dots, m\}$  es el vector de parámetros aleatorios con FDP conocida  $p_k(\vartheta)$ . La hipótesis  $H_0$  consiste en que la trayectoria  $\xi = (x_1, x_2, \dots, x_N)$  corresponde solo al ruido, mientras, la hipótesis alternativa  $H_1$  corresponde a la presencia del proceso aleatorio mas RBG. Hallar la relación de verosimilitud.



## 6.2. Estimación de parámetros en señales aleatorias

### 6.2.1. Estimación en condiciones de ruido aditivo

Sea una trayectoria de un proceso aleatorio,  $y(t) \in \mathfrak{Y}$ , que se observa en el intervalo  $(0, T)$ . Se asume, de forma a priori, que el proceso corresponde a una mezcla aditiva de una señal  $x(\boldsymbol{\vartheta}, t)$  con estructura determinística más una perturbación,  $\xi(t)$ :

$$y(t) = x(\boldsymbol{\vartheta}, t) + \xi(t) \quad (6.19)$$

donde  $\boldsymbol{\vartheta} = (\vartheta_1, \dots, \vartheta_m)$ ,  $\boldsymbol{\vartheta} \in \Theta$  es el vector de parámetros desconocidos de la señal, cuyo respectivo vector de estimación es un funcional de la trayectoria observada:

$$\tilde{\boldsymbol{\vartheta}} = \mathcal{K} \{y(t)\}$$

Por cierto,  $\tilde{\boldsymbol{\vartheta}} \in \Theta$ , en la medida en que usualmente el espacio de estimación coincide con el mismo espacio de los parámetros estimados. El valor medio de la perturbación se asume igual a cero,  $E\{\xi(t)\} = 0$ , y con función conocida de correlación propia,  $R_\xi(t, \tau)$ .

La estimación de los parámetros se puede realizar por cualquiera de los criterios de aproximación, descritos en el numeral §2.3. En el caso particular del empleo del criterio de máxima verosimilitud, dado el modelo (6.19), se obtiene el siguiente logaritmo para el funcional de verosimilitud:

$$\ln \Lambda(y(t)|\boldsymbol{\vartheta}) = \int_0^T v(t, \boldsymbol{\vartheta}) \left( y(t) - \frac{1}{2}x(t, \boldsymbol{\vartheta}) \right) dt \quad (6.20)$$

donde  $v(t, \boldsymbol{\vartheta})$  es la solución de la ecuación integral lineal no homogénea

$$\int_0^T R_\xi(t, \tau) v(\boldsymbol{\vartheta}, \tau) d\tau = x(t, \boldsymbol{\vartheta}), \quad 0 < t \in T \quad (6.21)$$

### 6.2.2. Estimación de máxima verosimilitud

La estimación, en este caso, se obtiene de la solución del respectivo sistema de ecuaciones (2.68), para lo cual inicialmente se determinan las derivadas parciales por cada uno de los parámetros del logaritmo del funcional de la relación de verosimilitud (6.20),

$$\begin{aligned} \frac{\partial}{\partial \vartheta_i} \ln \Lambda(y(t)|\boldsymbol{\vartheta}) &= \int_0^T \frac{\partial}{\partial \vartheta_i} v(t, \boldsymbol{\vartheta}) \left( y(t) - \frac{1}{2}x(t, \boldsymbol{\vartheta}) \right) dt \\ &\quad - \frac{1}{2} \int_0^T v(t, \boldsymbol{\vartheta}) \frac{\partial}{\partial \vartheta_i} x(t, \boldsymbol{\vartheta}) dt, \quad i = 1, \dots, m \end{aligned} \quad (6.22)$$

Sin embargo, de (6.21) se tiene que

$$\begin{aligned}
 \int_0^T v(t, \boldsymbol{\vartheta}) \frac{\partial}{\partial \vartheta_i} x(t, \boldsymbol{\vartheta}) dt &= \int_0^T v(t, \boldsymbol{\vartheta}) \left( \int_0^T R_\xi(t, u) \frac{\partial}{\partial \vartheta_i} v(u, \boldsymbol{\vartheta}) du \right) dt \\
 &= \int_0^T \frac{\partial}{\partial \vartheta_i} v(u, \boldsymbol{\vartheta}) \left( \int_0^T R_\xi(t, u) v(t, \boldsymbol{\vartheta}) dt \right) du \\
 &= \int_0^T \frac{\partial}{\partial \vartheta_i} v(u, \boldsymbol{\vartheta}) x(u, \boldsymbol{\vartheta}) du
 \end{aligned} \tag{6.23}$$

Reemplazando (6.23) en (6.22), entonces, se obtiene

$$\frac{\partial}{\partial \vartheta_i} \ln \Lambda(x(t)|\boldsymbol{\vartheta}) = \int_0^T \frac{\partial}{\partial \vartheta_i} v(u, \boldsymbol{\vartheta}) (y(t) - x(t, \boldsymbol{\vartheta})) dt, \quad i = 1, \dots, m \tag{6.24}$$

A partir de la expresión (6.24), se obtiene directamente el sistema de ecuaciones de máxima verosimilitud:

$$\int_0^T \frac{\partial}{\partial \vartheta_i} v(u, \boldsymbol{\vartheta}) (y(t) - x(t, \boldsymbol{\vartheta})) dt = 0, \quad i = 1, \dots, m \tag{6.25}$$

La solución para el sistema de ecuaciones de (6.25), en función de cada una de las variables  $(\vartheta_1, \dots, \vartheta_m)$ , corresponde a la estimación de máxima verosimilitud para el vector de parámetros  $\boldsymbol{\vartheta}$  de la señal,  $\tilde{\boldsymbol{\vartheta}} = (\tilde{\vartheta}_1, \dots, \tilde{\vartheta}_m)$ . La exigencia (2.69), sobre la consistencia de la estimación, impone una condición adicional por la cual la matriz informativa de Fisher:

$$\begin{aligned}
 I(\boldsymbol{\vartheta}; i, j) &= E \left\{ \frac{\partial}{\partial \vartheta_i} \ln \Lambda(y(t)|\boldsymbol{\vartheta}) \frac{\partial}{\partial \vartheta_j} \ln \Lambda(y(t)|\boldsymbol{\vartheta}) \right\} \\
 &= \int_0^T \int_0^T \frac{\partial}{\partial \vartheta_i} v(\tau_1, \boldsymbol{\vartheta}) \frac{\partial}{\partial \vartheta_j} v(\tau_2, \boldsymbol{\vartheta}) R_\xi(\tau_1, \tau_2) d\tau_1 d\tau_2, \quad i, j = 1, \dots, m
 \end{aligned}$$

debe ser positiva definida para  $\tilde{\boldsymbol{\vartheta}} = \boldsymbol{\vartheta}$ .

Cuando la perturbación  $\xi(t)$  corresponde al RBG, el cual tiene DEP igual a  $N_0/2$ , entonces de (6.25) se tiene que  $v(t, \boldsymbol{\vartheta}) = x(t, \boldsymbol{\vartheta})/N_0$ , con lo cual, el sistema de ecuaciones de máxima verosimilitud se simplifica hasta

$$\int_0^T \frac{\partial}{\partial \vartheta_i} x(t, \boldsymbol{\vartheta}) (y(t) - x(t, \boldsymbol{\vartheta})) dt = 0, \quad i = 1, \dots, m$$

En consecuencia, al ruido blanco Gaussiano le corresponden los siguientes elementos para

la matriz de Fisher:

$$I(\boldsymbol{\vartheta}; i, j) = \frac{1}{N_0} \int_0^T \frac{\partial}{\partial \vartheta_i} x(\tau, \boldsymbol{\vartheta}) \frac{\partial}{\partial \vartheta_j} x(\tau, \boldsymbol{\vartheta}) d\tau, \quad i, j = 1, \dots, m$$

### 6.2.3. Combinación lineal de señales

Sea la relación lineal entre los parámetros de la señal, donde  $x_i(t)$  son funciones conocidas, expresada como:

$$x(t, \boldsymbol{\vartheta}) = \sum_{j=1}^m \vartheta_j x_j(t), \quad (6.26)$$

La estimación conjunta de máxima verosimilitud para  $\vartheta_1, \dots, \vartheta_m$  se obtiene de reemplazar (6.26) en la parte derecha de (6.21), con lo cual se obtiene el sistema de ecuaciones:

$$\int_0^T R_\xi(t, \tau) v_i(\tau) d\tau = x_i(t), \quad i = 1, \dots, m \quad (6.27)$$

La función  $v(t, \boldsymbol{\vartheta})$ , de la que depende el logaritmo del funcional de verosimilitud, es:

$$v(t, \boldsymbol{\vartheta}) = \sum_{j=1}^m \vartheta_j v_j(t)$$

Por lo tanto,

$$\frac{\partial}{\partial \vartheta_i} v(t, \boldsymbol{\vartheta}) = v_i$$

que al reemplazar en (6.25), resulta en el sistema de ecuaciones de máxima verosimilitud,

$$\int_0^T v_i(t) \left( y(t) - \sum_{j=1}^m \vartheta_j x_j(t) \right) dt = 0, \quad i = 1, \dots, m \quad (6.28)$$

$$\text{luego, } \sum_{j=1}^m \vartheta_j \int_0^T v_i(t) x_j(t) dt = \int_0^T v_i(t) y(t) dt, \quad i = 1, \dots, m.$$

Al tomar las notaciones

$$x_{ij}(T) = \int_0^T v_i(t) x_j(t) dt \quad (6.29a)$$

$$y_i(T) = \int_0^T v_i(t) y(t) dt \quad (6.29b)$$

entonces, el sistema lineal de ecuaciones (6.28) se puede escribir en la forma:

$$\sum_{j=1}^m x_{ij}(T) \vartheta_j = y_i(T), \quad i = 1, \dots, m$$

que en forma matricial es

$$\mathbf{X}_{m \times m}(T) \boldsymbol{\vartheta}_{m \times 1} = \mathbf{y}_{1 \times m}(T) \quad (6.30)$$

Asumiendo que  $\int_0^T x_j^2(t) dt < \infty$ ,  $\forall j$ , además que  $R_\xi(t, \tau) > 0$ , entonces, se concluye que existe la matriz inversa  $\mathbf{X}^{-1}(T)$ , razón por la cual la solución de (6.30) conlleva a las siguientes estimaciones de máxima verosimilitud para los parámetros desconocidos:

$$\tilde{\boldsymbol{\vartheta}} = \mathbf{X}^{-1}(T) \mathbf{y}(T) \quad (6.31)$$

La comprobación sobre la consistencia de la estimación, se realiza empleando la parte izquierda de (6.28), de lo cual se encuentran los siguientes elementos informativos de la matriz de Fisher:

$$I(i, j) = E \left\{ \int_0^T v_i(\tau_1) \left( y(\tau_1) - \sum_{k=1}^m \vartheta_k x_k(\tau_1) \right) d\tau_1 \times \right. \\ \left. \int_0^T v_j(\tau_2) \left( y(\tau_2) - \sum_{n=1}^m \vartheta_n x_n(\tau_2) \right) d\tau_2 \right\}, \quad i, j = 1, \dots, m \quad (6.32)$$

De la expresión (6.32), se observa que la matriz informativa de Fisher es positiva definida, debido a que la función de correlación de la perturbación también es positiva definida. Por cierto, los elementos de la matriz informativa, en este caso, no dependen de los parámetros  $\vartheta_1, \dots, \vartheta_m$ . Por lo tanto, la solución (6.31) realmente representa una estimación consistente de máxima verosimilitud para el vector de parámetros  $\boldsymbol{\vartheta}$ . Además, las partes derechas de las expresiones (6.29a) y (6.32) coinciden, con lo cual la matriz informativa de Fisher  $I(T)$  es igual a la matriz  $\mathbf{X}(T)$ , entonces, la expresión (6.31) se puede representar en la forma

$$\tilde{\boldsymbol{\vartheta}} = I^{-1}(T) \mathbf{y}(T)$$

En el caso particular del RBG, al cual le corresponde una DEP igual a  $N_0$ , de (6.27) se obtiene que  $v_i(t) = x_i(t)/N_0$ ,  $i = 1, \dots, m$ , donde

$$x_{ij}(T) = \frac{1}{N_0} \int_0^T x_i(t) x_j(t) dt = I(i, j; T), \quad i, j = 1, \dots, m$$

$$y_i(T) = \frac{1}{N_0} \int_0^T x_i(t) y(t) dt, \quad i = 1, \dots, m$$

Cabe anotar que, en general, la estimación (6.31) no es centrada, en la medida en que

$$E \{ \tilde{\boldsymbol{\theta}} \} = \mathbf{X}^{-1}(T) E \{ \mathbf{y}(T) \} \quad (6.33)$$

Sin embargo,

$$E \{ y_i(T) \} = \sum_{j=1}^m \vartheta_j \int_0^T x_j(t) v_i(t) dt = \sum_{j=1}^m \vartheta_j x_{ij}(T)$$

entonces

$$E \{ \mathbf{X}(T) \} = \mathbf{y}(T) \boldsymbol{\vartheta} \quad (6.34)$$

Reemplazando (6.34) en (6.33), finalmente se obtiene que  $E \{ \tilde{\boldsymbol{\theta}} \} = \boldsymbol{\vartheta}$  lo que demuestra la ausencia de sesgo en la estimación (6.31).

**Ejemplo 6.2.** Hallar la estimación de máxima verosimilitud para una señal con amplitud desconocida,  $\mathbf{ax}(t)$ , que está inmersa en ruido blanco Gaussiano.

La señal en análisis se considera un caso particular de (6.26), cuando  $m = 1$  y  $\vartheta_1 = a$ . Luego, la estimación para el respectivo modelo lineal de señal en (6.31) toma la forma:

$$\tilde{a} = \int_0^T v(t) y(t) dt \Big/ \int_0^T v(t) x(t) dt \quad (6.35)$$

donde

$$y(T) = \int_0^T v(t) y(t) dt, \quad x(T) = \int_0^T v(t) x(t) dt$$

siendo  $v(t)$  la solución de la ecuación lineal integral dada en (6.27),

$$\int_0^T R_{\xi}(t, \tau) v(\tau) d\tau = x(t), \quad 0 \leq t \leq T$$

Como antes se indicó, la estimación (6.35) de máxima verosimilitud para un parámetro del modelo lineal de la señal no tiene sesgo y es efectiva, esto es

$$E \{ \tilde{a} \} = a$$

$$\mu_2 \{ \tilde{a} \} = \mathbf{X}_T^{-1} = I^{-1}(T) = \left( \int_0^T v(t) x(t) dt \right)^{-1}$$

Al asumir el modelo del ruido blanco Gaussiano, entonces,  $v(t) = x(t)/N_0$ , y de (6.35) se

tiene, entonces, que

$$\tilde{a} = \int_0^T x(t)y(t)dt \Big/ \int_0^T x^2(t)dt$$

La implementación del algoritmo de máxima verosimilitud para la estimación de la amplitud de la señal, cuando se tiene RBG, exige el cálculo de la integral normalizada de correlación (6.35). Esta operación se realiza mediante la filtración lineal con respuesta a impulso [12]:

$$h(\tau) = \begin{cases} x(T - \tau) \Big/ \int_0^T x^2(t)dt, & 0 \leq \tau \leq T \\ 0, & \tau < 0, \tau > T \end{cases}$$

### Problemas

**Problema 6.2.** Dado el modelo de combinación lineal de señales (6.26),

- Demostrar que la estimación vectorial de máxima verosimilitud (6.31) no tiene sesgo.
- Hallar la matriz de correlación de la estimación de máximo verosimilitud  $E\{(\tilde{\boldsymbol{\vartheta}} - \boldsymbol{\vartheta})(\tilde{\boldsymbol{\vartheta}} - \boldsymbol{\vartheta})^T\}$  y demostrar la efectividad de la estimación (6.31).

**Problema 6.3.** Asumir que se tiene una trayectoria  $y(t)$  de la suma de señal y RBG (6.31), observada en el intervalo de tiempo  $(0, T)$ , la cual es discretizada para obtener la correspondiente sucesión  $\mathbf{y} = \{y(t_i) : i = 1, \dots, n\}$ ,  $t_i \in T$ :

- Hallar la correspondiente estimación de máximo verosimilitud.
- Demostrar que la estimación obtenida es insesgada y hallar su matriz de correlación.

**Problema 6.4.** Al igual que en el problema anterior, asumir la discretización de la trayectoria de la señal  $ax(t)$  del ejemplo 6.2, para la estimación de la amplitud desconocida  $a$ . Hallar la estimación de máximo verosimilitud.

### 6.3. Filtración óptima lineal por mínimos cuadrados

En este caso, la filtración optimiza la forma de representación de los dispositivos lineales de proceso, como observa en la Figura 6.2, en la cual las observaciones iniciales  $x$  pertenecientes al proceso  $\xi$  interactúan con la perturbación  $\eta$ , en forma general se analiza la combinación lineal; acción que se representa mediante el funcional de transformación  $\mathcal{K}_1\{\xi, \eta\}$ , cuya señal de salida pasa por el filtro lineal, representado por la respuesta a impulso  $h(t, s)$  y que se escoge, de tal manera que la salida del filtro en el momento  $t$  corresponda al mínimo del error cuadrático medio (2.64) de la estimación  $\theta \subset x$ , que se contiene en la señal  $\xi$ . El mismo parámetro en estimación se puede describir por la transformación  $\mathcal{K}_0\{\xi\}$ .

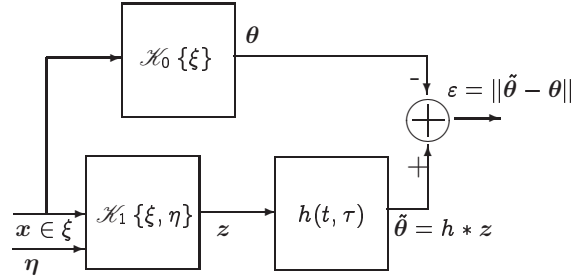


Figura 6.2. Filtración de señales

Cuando se analiza la transmisión de señales, durante la cual es importante asegurar el mínimo de distorsión, el operador de transformación  $\mathcal{K}_0$  describe un canal ideal, y por lo tanto, se hace necesario sintetizar un filtro con respuesta a impulso, conectado en serie al canal real con operador de transformación  $\mathcal{K}_1$ , de tal manera que su salida corresponda a alguna transformación tan cercana como se pueda a  $\mathcal{K}_0$ .

#### 6.3.1. Optimización de la respuesta a impulso

Sea  $z(t)$  una trayectoria para la combinación lineal en forma de proceso aleatorio aditivo:  $z(t) = x(t) + \eta(t)$ , tal que  $\|x\|, \|\eta\| < \infty$ . En calidad de estimación  $\tilde{\theta}(t)$  se toma el valor filtrado de la observación:

$$\tilde{\theta}(t) = \int_{-\infty}^{\infty} h(t, \tau) z(\tau) d\tau \quad (6.36)$$

siendo  $h(t, \tau)$  la respuesta a impulso del dispositivo lineal, que por ahora se asume sin restricciones de causalidad en su implementación práctica. No obstante, el filtro se considera que cumple con la condición de estabilidad.

La estimación (6.36) es ante todo sesgada, en la medida en que

$$m_{1\tilde{\theta}}(t) = \int_{-\infty}^{\infty} h(t, \tau) (m_{1x}(\tau) + m_{1\eta}(\tau)) d\tau \neq m_{1x}(t)$$

donde  $m_{1\eta}(t)$  y  $m_{1x}(t)$  son las respectivas medias de la señal útil y de la perturbación.

El valor cuadrático medio del error, determinado como  $\varepsilon = (\theta - \tilde{\theta})^2$ , se obtiene mediante

su promedio de tiempo, que en caso emplear la estimación (6.36), es igual a:

$$\varepsilon^2(t) = E \left\{ \left( \theta(t) - \tilde{\theta}(t) \right)^2 \right\} \quad (6.37)$$

Al tomar el error, (6.37), como criterio de calidad se puede encontrar la respuesta a impulso que brinde su mínimo valor con respecto a todos los demás posibles sistemas lineales. En particular, al reemplazar (6.36) en (6.37), se obtiene:

$$\begin{aligned} \varepsilon^2(t) &= E \left\{ \theta^2(t) - 2 \int_{-\infty}^{\infty} h(t, \tau) z(\tau) \theta(t) d\tau \right. \\ &\quad \left. + \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h(t, \lambda_1) h(t, \lambda_2) z(\lambda_1) z(\lambda_2) d\lambda_1 d\lambda_2 \right\} \\ &= E \left\{ \theta^2(t) \right\} - 2 \int_{-\infty}^{\infty} h(t, \tau) E \{ z(\tau) \theta(t) \} d\tau + \\ &\quad + \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h(t, \lambda_1) h(t, \lambda_2) E \{ z(\lambda_1) z(\lambda_2) \} d\lambda_1 d\lambda_2 \end{aligned} \quad (6.38)$$

Sean conocidas las funciones de correlación propia,  $R_i(t_1, t_2)$ ,  $i = \{z, \theta\}$ , además, de la respectiva función de correlación mutua  $R_{z\theta}(t_1, t_2)$  entre los procesos  $z(t)$  y  $\theta(t)$ .

Por cuanto,

$$\begin{aligned} E \{ z(\lambda_1) z(\lambda_2) \} &= R_z(\lambda_1, \lambda_2) \\ &= R_x(\lambda_1, \lambda_2) + R_\xi(\lambda_1, \lambda_2) + R_{\xi x}(\lambda_1, \lambda_2) + R_{x\xi}(\lambda_1, \lambda_2) \end{aligned} \quad (6.39a)$$

$$E \{ z(\tau) \theta(t) \} = R_{z\theta}(\lambda_1, \lambda_2) = R_\theta(\tau, t) + R_{z\theta}(\tau, t) \quad (6.39b)$$

entonces, reemplazando (6.39a) y (6.39b) en (6.38) se obtiene que

$$\begin{aligned} \varepsilon^2(t) &= R_\theta(t, t) - 2 \int_{-\infty}^{\infty} h(t, \tau) R_{z\theta}(\tau, t) d\tau \\ &\quad + \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h(t, \lambda_1) h(t, \lambda_2) R_z(\lambda_1, \lambda_2) d\lambda_1 d\lambda_2 \end{aligned} \quad (6.40)$$

La expresión (6.40) muestra que el error cuadrático medio de la estimación lineal  $\tilde{\theta}(t)$  depende, tanto de las funciones de correlación propia  $R_i(t_1, t_2)$ ,  $i = \{z, \theta\}$ , como de la función de correlación mutua  $R_{z\theta}(t_1, t_2)$  de los procesos  $z(t)$  y  $\theta(t)$ , pero de ninguna manera de la estructura más fina de estos procesos [12]. En general, se puede demostrar que, dadas todas las anteriores funciones de correlación, la mejor estimación lineal en el sentido del error cuadrático medio corresponde al filtro con respuesta a impulso que cumpla



la ecuación integral [43]:

$$R_{z\theta}(\tau, t) = \int_{-\infty}^{\infty} h_{opt}(t, \lambda) R_z(\tau, \lambda) d\lambda \quad (6.41)$$

El reemplazo de (6.41) en (6.40), resulta en

$$\begin{aligned} \varepsilon^2(t) = R_\theta(t, t) - 2 \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h(t, \tau) h_{opt}(t, \lambda) R_z(\tau, \lambda) d\tau d\lambda \\ + \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h(t, \lambda_1) h(t, \lambda_2) R_z(\lambda_1, \lambda_2) d\lambda_1 d\lambda_2 \end{aligned}$$

luego

$$\begin{aligned} \varepsilon^2(t) = R_\theta(t, t) - \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h_{opt}(t, \tau) h_{opt}(t, \lambda) R_z(\tau, \lambda) d\tau d\lambda \\ + \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} R_z(\lambda_1, \lambda_2) \{h(t, \lambda_1) - h_{opt}(t, \lambda_1)\} \{h(t, \lambda_2) - h_{opt}(t, \lambda_2)\} d\lambda_1 d\lambda_2 \end{aligned} \quad (6.42)$$

Por cuanto, sólo el último término de (6.42) contiene la función desconocida  $h(t, \lambda)$ , la cual debe ser definida positiva [12], entonces el menor valor de  $\overline{\varepsilon^2(t)}$  ocurre cuando el último término es 0, esto es, cuando el filtro tiene respuesta a impulso  $h(t, \lambda) \equiv h_{opt}(t, \lambda)$ . En este caso, el valor mínimo de error será:

$$\varepsilon_{\min}^2(t) = R_\theta(t, t) - \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h_{opt}(t, \tau) h_{opt}(t, \lambda) R_z(\tau, \lambda) d\tau d\lambda$$

Teniendo en cuenta (6.41), entonces

$$\varepsilon_{\min}^2(t) = R_\theta(t, t) - \int_{-\infty}^{\infty} h_{opt}(t, \lambda) R_{z\theta}(\lambda, t) d\lambda$$

De otra parte, la expresión (5.8) relaciona las funciones de correlación propia a la entrada y salida de un sistema lineal, con lo cual, la última expresión toma la forma definitiva,

$$\varepsilon_{\min}^2(t) = R_\theta(t, t) - R_{\hat{\theta}}(t, t) \quad (6.43)$$

la cual implica, en primer orden, que el mínimo error cuadrático medio (2.64) corresponde a la diferencia de los valores cuadráticos medios del proceso a estimar y de su estimación. En segundo lugar, se tiene que  $R_{\hat{\theta}}(t, t) \leq R_\theta(t, t)$ .

Si se asume la estacionariedad de los procesos  $\theta(t)$  y  $\eta(t)$ , por lo menos en el sentido amplio, además el filtro se supone con parámetros en (1.38) invariables en el tiempo, entonces la ecuación integral (6.41) toma la forma denominada *ecuación de Wiener-Hopf*:

$$R_{z\theta}(\tau) = \int_{-\infty}^{\infty} h_{opt}(\lambda) R_z(\tau - \lambda) d\lambda \quad (6.44)$$

La invariabilidad de los momentos de aleatoriedad de los procesos estacionarios, supone la invariabilidad en el valor del error cuadrático medio, cuyo valor mínimo (6.43), es:

$$\begin{aligned} \varepsilon_{\min}^2 &= R_{\theta}(0) - \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h_{opt}(\lambda_1) h_{opt}(\lambda_2) R_z(\lambda_1 - \lambda_2) d\lambda_1 d\lambda_2 \\ &= R_{\theta}(0) - \int_{-\infty}^{\infty} h_{opt}(\lambda) R_{z\theta}(\lambda) d\lambda \end{aligned} \quad (6.45)$$

con lo que el error cuadrático medio es igual a

$$\varepsilon_{\min}^2 = R_{\theta}(0) - R_{\hat{\theta}}(0) = m_{2\theta} - m_{2\hat{\theta}} \quad (6.46)$$

esto es, se determina por la diferencia de los valores medios de potencia, tanto del proceso a estimar, como de la propia estimación.

En concordancia con (3.24), los valores medios de potencia se pueden expresar a través de las respectivas densidades espectrales de potencia,

$$\varepsilon_{\min}^2 = \frac{1}{2\pi} \int_0^{\infty} \left( S_{\theta}(\omega) - S_z(\omega) |H_{opt}(\omega)|^2 \right) d\omega \quad (6.47)$$

siendo  $H_{opt}(\omega)$  la función de transferencia del filtro óptimo lineal, además

$$S_z(\omega) = S_{\theta}(\omega) + S_{\eta}(\omega) + S_{\theta\eta}(\omega) + S_{\eta\theta}(\omega) \quad (6.48)$$

donde  $S_{\theta}(\omega)$ ,  $S_{\eta}(\omega)$ ,  $S_{\theta\eta}(\omega)$  y  $S_{\eta\theta}(\omega)$  son las densidades espectrales de potencia propias y mutuas de los procesos  $\theta(t)$  y  $\eta(t)$ , de manera correspondiente.

De esta manera, la respuesta a impulso óptima de un sistema lineal, en caso de asumir la estacionariedad de los procesos, consiste en la solución de la ecuación integral (6.44), sin tener en cuenta la condición de causalidad, mediante la TF en la forma,

$$S_{z\theta}(\omega) = S_{\theta}(\omega) + S_{z\theta}(\omega) = H_{opt}(\omega) S_z(\omega)$$

de donde, teniendo en cuenta (6.48), se tiene

$$H_{opt}(\omega) = \frac{S_{z\theta}(\omega)}{S_z(\omega)} = \frac{S_{\theta}(\omega) + S_{z\theta}(\omega)}{S_{\theta}(\omega) + S_{\eta}(\omega) + S_{\theta\eta}(\omega) + S_{\eta\theta}(\omega)} \quad (6.49)$$

Reemplazando (6.49) en (6.47) se obtiene el valor del mínimo error cuadrático medio,

$$\varepsilon_{\min}^2 = \frac{1}{2\pi} \int_0^\infty \frac{S_\theta(\omega) S_z(\omega) - |S_{z\theta}(\omega)|^2}{S_z(\omega)} d\omega \quad (6.50)$$

Si se considera la correlación nula entre las señales  $\theta(t)$  y  $\eta(t)$ , esto es,  $R_{\theta\eta}(\tau) = 0$ , entonces, sus DEP mutuas también son nulas,  $S_{\theta\eta}(\omega) = S_{\eta\theta}(\omega) = 0$ , por lo que la función de transferencia óptima de un sistema lineal tiene la forma,

$$H_{opt}(\omega) = \frac{S_\theta(\omega)}{S_\theta(\omega) + S_\eta(\omega)}$$

De (6.50), y reemplazando  $S_{z\theta}(\omega) = S_\theta(\omega)$ , además  $S_z(\omega) = S_\theta(\omega) + S_\eta(\omega)$ , se obtiene para las señales de correlación nula,

$$\varepsilon_{\min}^2 = \frac{1}{2\pi} \int_0^\infty \frac{S_\theta(\omega) S_\eta(\omega)}{S_\theta(\omega) + S_\eta(\omega)} d\omega \quad (6.51)$$

El error cuadrático medio (6.51) se puede hacer igual a cero, cuando las densidades espectrales de potencia de las señales  $\theta(t)$  y  $\eta(t)$  no se traslapan, esto es, cuando para todo valor de  $\omega$  se cumple que  $S_\eta(\omega) = S_\theta(\omega) = 0$ , lo cual a su vez implica que por lo menos uno de los dos espectros debe tener ancho de banda finito. En caso contrario, siempre ocurrirá un error.

En síntesis, la filtración de una señal estacionaria a partir de su combinación lineal con otro proceso estacionario, de tal manera que el error cuadrático medio sea 0, corresponde al caso en que las respectivas DEP de ambos procesos tienen rangos de frecuencia en los cuales no hay aporte energético. En este sentido, si se analiza la filtración del proceso  $\theta(t)$  dado sobre un fondo de ruido blanco Gaussiano, entonces de (6.51) se obtiene

$$\varepsilon_{\min}^2 = \frac{N_0}{2\pi} \int_0^\infty \frac{S_\theta(\omega)}{S_\theta(\omega) + N_0} d\omega$$

que implica, que siempre se tiene en condiciones reales un valor diferente de 0 para el error cuadrático medio.

### 6.3.2. Condición de realización física

Un filtro lineal con respuesta a impulso óptima (6.41) no es físicamente realizable, al no cumplir la condición de causalidad, por cuanto  $h_{opt}(t, \lambda) \neq 0$ , para  $t < \lambda$ . Al observar en el numeral §6.3.1, la operación de filtración se realiza después de analizada la combinación lineal de las señales  $\theta(t)$  y  $\eta(t)$  sobre todo el intervalo de tiempo  $t \in (-\infty, \infty)$ , lo cual implica que la filtración óptima realiza la estimación del valor del proceso en un momento dado del tiempo con retardo infinito.

La condición de realización física implica que la filtración debe realizarse sobre la observación  $z(t)$  determinada hasta el momento de tiempo, en el cual se realiza la estimación,

$$h(t, \tau) = 0, \quad \forall \tau < 0 \quad (6.52)$$

Reemplazando la condición de causalidad (6.52) en (6.36) se obtiene la estimación de  $\theta(t)$  con un filtro físicamente realizable,

$$\tilde{\theta}(t) = \int_{-\infty}^t h(t, \tau) z(\tau) d\tau$$

Frecuentemente, la estimación se realiza sobre intervalos cerrados y finitos de análisis,  $(t - T, t)$ , por lo que para la anterior integral se ajustan los límites en la siguiente forma,

$$\tilde{\theta}(t) = \int_{t-T}^t h(t, \tau) z(\tau) d\tau = \int_0^T h(t, t - \lambda) z(t - \lambda) d\lambda \quad (6.53)$$

El reemplazo de (6.53) en (6.37), y al efectuar las mismas transformaciones de (6.38) se obtiene la correspondiente expresión del error cuadrático medio del filtro óptimo que se ajusta a la condición de realización física, asumiendo un intervalo finito de estimación:

$$\begin{aligned} \varepsilon^2(t) = & R_{\theta}(t, t) - 2 \int_0^T h(t, t - \tau) R_{z\theta}(t - \tau, t) d\tau + \\ & + \int_0^T \int_0^T h(t, t - \lambda_1) h(t, t - \lambda_2) R_z(t - \lambda_1, t - \lambda_2) d\lambda_1 d\lambda_2 \end{aligned} \quad (6.54)$$

El filtro lineal óptimo para la definición del error acotado (6.54), implica que su respuesta impulso causal debe cumplir la ecuación integral,

$$R_{z\theta}(t - \tau, t) = \int_T^t h_{opt}(t, t - \lambda) R_z(t - \tau, t - \lambda) d\lambda, \quad \tau \in T \quad (6.55)$$

con lo cual el valor mínimo del error cuadrático medio del filtro óptimo físicamente implementable es igual a

$$\varepsilon_{\min}^2(t) = R_{\theta}(t, t) - \int_0^T h_{opt}(t, t - \lambda) R_{z\theta}(t - \lambda, t) d\lambda$$

Al asumir la estacionariedad de los procesos  $\theta(t)$  y  $\eta(t)$ , además suponiendo la invariabilidad de sus parámetros en el tiempo, entonces, las anteriores dos relaciones toman la

siguiente forma definitiva:

$$R_{z\theta}(\tau) = \int_T h_{opt}(\lambda) R_z(\tau - \lambda) d\lambda, \tau \in T \quad (6.56a)$$

$$\varepsilon_{\min}^2 = R_\theta(0) - \int_T h_{opt}(\lambda) R_{z\theta}(\lambda) d\lambda \quad (6.56b)$$

**Ejemplo 6.3.** Sea la señal  $\theta = as(t)$ , siendo  $s(t)$  una función conocida  $a$  y la amplitud aleatoria. Hallar la respuesta del filtro lineal que minimize el error cuadrático medio de la estimación  $\tilde{\theta} = \tilde{a}s(t)$  hecha de la trayectoria  $\xi(t) + \eta(t) \in (t-T, t)$ , siendo  $\eta(t)$  un proceso aleatorio independiente de  $a$ , con  $\overline{\eta(t)} = 0$  y función de correlación  $R_\eta(t_1, t_2)$ .

La solución de la ecuación integral (6.55), reemplazando  $R_{z\theta}(t_1, t_2)$  y  $R_\theta(t_1, t_2)$  por:

$$\begin{aligned} R_{z\theta}(t - \tau, t) &= E\{a^2\} s(t - \tau) s(t) \\ R_z(t - \tau, t - \lambda) &= E\{a^2\} s(t - \tau) s(t - \lambda) + R_\eta(t - \tau, t - \lambda) \end{aligned}$$

con lo cual, para  $\tau \in T$  se obtiene que

$$E\{a^2\} s(t - \tau) s(t) = \int_0^T h_{opt}(t, t - \lambda) (E\{a^2\} s(t - \tau) s(t - \lambda) - R_\eta(t - \tau, t - \lambda)) d\lambda,$$

La solución de la anterior ecuación integral tiene la forma,

$$h_{opt}(t, \tau) = ks(t)v(t)$$

siendo  $v(t)$  la solución de la ecuación integral [44],

$$\int_{t-T}^t v(\lambda_1) R_\eta(\lambda_2, \lambda_1) d\lambda_1 = s(\lambda_2), \lambda_2 \in (t-T, t)$$

Al reemplazar las anteriores expresiones en la solución buscada de (6.55) se obtiene,

$$E\{a^2\} = k \left( E\{a^2\} \int_0^T v(t - \lambda) s(t - \lambda) d\lambda + 1 \right)$$

de donde se encuentra el siguiente valor de  $k$ :

$$k = E\{a^2\} \left( E\{a^2\} \int_0^T v(t - \lambda) s(t - \lambda) d\lambda + 1 \right)^{-1}$$

El filtro óptimo, en este caso de estimación, tiene respuesta a impulso,

$$h_{opt}(t, \tau) = \frac{E\{a^2\} s(t) v(t)}{1 + E\{a^2\} \int_{t-T}^t v(\lambda) s(\lambda) d\lambda}$$

A partir de la trayectoria  $z(t)$ , que implica la adición del proceso  $\eta(t)$ , observada en el intervalo de tiempo  $(t-T, t)$ , la estimación de la señal en el momento  $t$  tiene la forma:

$$\tilde{a}s(t) = \int_0^T h_{opt}(t, t-\lambda) z(t-\lambda) d\lambda = \frac{E\{a^2\} s(t) \int_{t-T}^t z(\lambda) v(\lambda) d\lambda}{1 + E\{a^2\} \int_{t-T}^t v(\lambda) s(\lambda) d\lambda}$$

por lo cual, la estimación de la misma amplitud de la señal es igual a

$$\tilde{a} = \frac{E\{a^2\} \int_{t-T}^t z(\lambda) v(\lambda) d\lambda}{1 + E\{a^2\} \int_{t-T}^t v(\lambda) s(\lambda) d\lambda}$$

que resulta ser sesgada, como se observa de su valor promedio,

$$E\{\tilde{a}\} = E\{a\} \frac{E\{a^2\} \int_{t-T}^t s(\lambda) v(\lambda) d\lambda}{1 + E\{a^2\} \int_{t-T}^t v(\lambda) s(\lambda) d\lambda}$$

Aunque asintóticamente, cuando  $E\{a^2\} \rightarrow \infty$ , esta puede ser considerada como no sesgada. El respectivo error cuadrático medio de la estimación  $\tilde{a}$  está dado por la expresión:

$$\begin{aligned} \varepsilon_{\min}^2(t) &= E\{a^2\} s^2(t) - k E\{a^2\} \int_0^T s^2(t) s(t-\tau) v(t-\tau) d\tau \\ &= E\{a^2\} s^2(t) \left( 1 - k \int_{t-T}^t v(\lambda) s(\lambda) d\lambda \right) \end{aligned}$$

que al reemplazar el valor de la constante  $k$  resulta en

$$\varepsilon_{\min}^2(t) = E\{(\tilde{a} - a)^2 s^2(t)\} = \frac{E\{a^2\} s^2(t)}{1 + E\{a^2\} \int_{t-T}^t v(\lambda) s(\lambda) d\lambda}$$

En el caso particular, cuando la filtración del parámetro aleatorio se considera inmersa en ruido blanco aditivo con densidad espectral de potencia  $N_0/2$  y función de correlación

$$R_\eta(t, \lambda) = \frac{N_0}{2} \delta(t - \lambda), v(\tau) = \frac{2}{N_0} s(\tau)$$

con lo cual, las correspondientes respuesta a impulso óptima y valor de error cuadrático medio tiene la forma:

$$h_{opt}(t, \tau) = \frac{\frac{2E\{a^2\}}{N_0} s(t) s(\tau)}{1 + \frac{2E\{a^2\}}{N_0} \int_{t-T}^t s^2(\lambda) d\lambda}$$

$$\varepsilon_{\min}^2(t) = \frac{E\{a^2\} s^2(t)}{1 + \frac{2E\{a^2\}}{N_0} \int_{t-T}^t s^2(\lambda) d\lambda}$$

### 6.3.3. Filtros acoplados

En el numeral §6.3, se analiza la calidad de filtración mediante el criterio del error cuadrático medio, que se justifica cuando en el proceso de la señal tiene sentido el análisis de la dependencia del error en el tiempo. En otras tareas, es más importante establecer simplemente la presencia o ausencia de la señal inmersa en la perturbación, lo que puede ser resuelto mediante la filtración que brinde la máxima relación señal/ruido, sin importar cuánto se afecte la señal útil.

Sea  $z(t) = x(t) + \eta(t)$ , la combinación lineal de la señal útil y la perturbación aleatoria  $\eta(t)$ , que se supone estacionaria (en el sentido amplio) con valor medio  $\overline{\eta(t)} = 0$  y función conocida de correlación propia  $R_\eta(\tau)$ . La estimación de la señal en el intervalo de observación  $T$  corresponde al valor

$$\tilde{x}(t) = \int_0^T h(\tau) z(t - \tau) d\tau \quad (6.57)$$

donde  $h(\tau)$  es la función respuesta a impulso que debe ser determinada, asumiendo la linealidad del sistema y la invariabilidad de sus parámetros de definición en el tiempo. La relación anterior se puede escribir de la forma:

$$\tilde{x}(t) = x_1(t) + v(t) \quad (6.58)$$

donde

$$x_1(t) = \int_0^T h(\tau) x(t - \tau) d\tau, \quad v(t) = \int_0^T h(\tau) \eta(t - \tau) d\tau$$

La relación por potencia de señal/ruido se determina como la relación del valor cuadrático medio la señal de salida  $\overline{x_1^2(t)}$ , estimada en un momento determinado del tiempo  $t_0$ , sobre

la varianza del ruido a la salida del filtro  $\sigma_v^2$ , así,

$$S/N = \frac{\overline{x_1^2(t)}}{\sigma_v^2} \quad (6.59)$$

El filtro lineal óptimo, en el sentido del máximo valor de la relación  $S/N \triangleq \mu_{\max}$ , se denomina *filtro acoplado*.

En cualquier sistema lineal con respuesta a impulso  $h(\tau)$ , teniendo en cuenta (6.59), se cumple la siguiente desigualdad:

$$\begin{aligned} \mu_{\max} \sigma_v^2 - x_1^2(t_0) &= \mu_{\max} \int_0^T \int_0^T h(\lambda_1) h(\lambda_2) R_\eta(\lambda_2 - \lambda_1) d\lambda_2 d\lambda_1 - \\ &\quad - \left( \int_0^T h(\tau) x(t_0 - \tau) d\tau \right)^2 \geq 0 \end{aligned} \quad (6.60)$$

de tal manera, que el símbolo de la igualdad, que corresponde al caso del filtro acoplado, ocurre para la función  $h(\tau) = h_{ac}(\tau, t_0)$ , que cumpla con la ecuación integral:

$$\int_0^T h_{ac}(\tau; t_0) R_\eta(t - \tau) d\tau = k(t_0) x(t_0 - t), \quad t \in [0, T] \quad (6.61)$$

siendo  $k(t_0)$  una constante, que para el momento  $t_0$  es igual a

$$k(t_0) = \frac{\int_0^T h_{ac}(\tau; t_0) x(t_0 - \tau) d\tau}{\mu_{\max}} = \frac{x_{1c}(t_0)}{\mu_{\max}} = \frac{\sigma_{vc}^2}{x_{1c}(t_0)}$$

donde la señal  $x_{1c}$  y la varianza  $\sigma_{vc}^2$  están dados a la salida del filtro acoplado. Cabe anotar, que si existe la función  $h_{ac}(\tau, t_0)$  como solución de (6.61), entonces esta misma ecuación se cumple para  $c h_{ac}(\tau, t_0)$ , siendo  $c$  una constante (por cierto, el valor  $\mu_{\max}$  no cambia), con lo cual la respuesta a impulso del filtro acoplado puede ser determinada hasta el valor de su factor de escala.

Una solución directa se puede hallar al asumir que la perturbación corresponde a ruido blanco Gaussiano con función de correlación propia  $R_{eta}(t - \tau) = N_0 \delta(t - \tau)$ . Al despejar la función delta dentro de la integral, mediante la propiedad de selectividad, la respuesta a impulso del filtro acoplado toma la forma,

$$h_{ac}(t; t_0) = \frac{k(t_0)}{N_0} x(t_0 - t), \quad t \in [0, T] \quad (6.62)$$

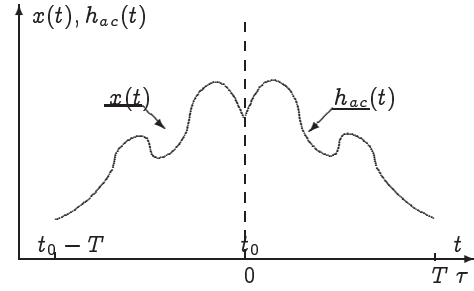
El análisis de la estructura de  $h_{ac}(t; t_0)$  en (6.62), muestra que la respuesta a impulso del filtro acoplado, obtenida en este caso, es directamente proporcional a la forma imagen



con respecto al eje vertical, que pasa por el valor de tiempo  $t = t_0$ , de la señal útil  $x(t)$  en el intervalo  $(t_0 - T, t_0)$ , con el consecuente traslado del origen de las coordenadas hasta el punto  $t = t_0$ . En el caso simple cuando  $t_0 = T$ , la respuesta a impulso del filtro acoplado corresponde a la representación espejo de la señal con respecto al eje vertical, que divide en dos el intervalo de análisis  $(0, T)$ , como se observa en la Figura 6.3.

El máximo valor de la relación señal/ruido, a partir de (6.60), es igual a

$$\begin{aligned}\mu_{\max} &= \frac{\left( \int_0^T x^2(t_0 - \tau) d\tau \right)^2}{N_0 \int_0^T x^2(t_0 - \tau) d\tau} \\ &= \frac{1}{N_0} \int_0^T x^2(t_0 - \tau) d\tau = \frac{\mathcal{E}_x}{N_0}\end{aligned}$$



**Figura 6.3.** Respuesta a impulso de un filtro acoplado

esto es, el valor corresponde a la relación de energía de la señal sobre la DEP del ruido  $N_0$ , en el intervalo  $(t_0 - T, t_0)$ .

La función de transferencia  $H_{ac}(\omega; t_0)$  del filtro acoplado con respuesta a impulso (6.62), se determina como

$$\begin{aligned}H_{ac}(\omega; t_0) &= \int_{-\infty}^{\infty} h_{ac}(t; t_0) e^{-j\omega t} dt = \frac{k(t_0)}{N_0} \int_0^T x(t_0 - t) e^{-j\omega t} dt \\ &= \frac{k(t_0)}{N_0} e^{-j\omega t_0} \int_{t_0 - T}^{t_0} x(t) e^{j\omega t} dt\end{aligned}$$

lo que puede ser interpretado como

$$H_{ac}(\omega; t_0) = \frac{k(t_0)}{N_0} e^{-j\omega t_0} X^*(\omega; t_0, T)$$

Por lo anterior, durante la separación de una señal de su mezcla aditiva con el ruido blanco, la respuesta a impulso del filtro acoplado es proporcional al espectro conjugado  $X^*(\omega; t_0, T)$  de la señal útil y del truncado en el intervalo  $(t_0 - T, t_0)$ .

En forma general, se puede obtener la expresión de la respuesta a impulso para el caso de cualquier DEP de ruido dado  $\mathcal{S}_\eta(\omega)$ , pero sin tener en cuenta la condición para la realización física del filtro, de manera particular, al suponer que se tiene la observación

$z(t)$  para todos los valores de  $t \in \mathbb{R}$ , generalizando los límites de la integral (6.57):

$$\tilde{x}(t) = \int_{-\infty}^{\infty} h(\tau; t_0) z(t - \tau) d\tau$$

En este caso, la ecuación (6.60), para todos los valores de  $t$ , toma la siguiente forma:

$$\int_{-\infty}^{\infty} h_{ac}(\tau; t_0) R_{\eta}(t - \tau) d\tau = k(t_0) x(t_0 - t)$$

que al hallar la transformada de Fourier se convierte en:

$$H_{ac}(\omega; t_0) S_{\eta}(\omega) = k(t_0) X^*(\omega) e^{j\omega t_0}$$

con lo cual, la respuesta a impulso del filtro acoplado generalizado tiene la forma:

$$H_{ac}(\omega; t_0) = k(t_0) \frac{X^*(\omega)}{S_{\eta}(\omega)} e^{-j\omega t_0} \quad (6.63)$$

El máximo valor de la relación señal/ruido, en concordancia con (6.60) se calcula como:

$$\begin{aligned} \mu_{\max} &= \frac{\left( \int_{-\infty}^{\infty} h_{ac}(\tau, t_0) x(t_0 - \tau) d\tau \right)^2}{\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h_{ac}(\lambda_1; t_0) h_{ac}(\lambda_2; t_0) R_{\eta}(\lambda_2 - \lambda_1) d\lambda_1 d\lambda_2} \\ &= \frac{\left( \frac{1}{2\pi} \int_{-\infty}^{\infty} H_{ac}(\omega; t_0) X(\omega) e^{j\omega t} d\omega \right)^2}{\frac{1}{2\pi} \int_{-\infty}^{\infty} S_{\eta}(\omega) |H_{ac}(\omega; t_0)|^2 d\omega} \end{aligned}$$

valor en el cual, al sustituir  $H_{ac}(\omega; t_0)$  por (6.63), se obtiene la expresión final

$$\mu_{\max} = \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{|X(\omega)|^2}{S_{\eta}(\omega)} d\omega \quad (6.64)$$

Cuando la perturbación corresponde al ruido blanco Gaussiano, de la expresión (6.64) resulta entonces el siguiente valor

$$\mu_{\max} = \frac{1}{2\pi N_0} \int_{-\infty}^{\infty} |X(\omega)|^2 d\omega$$

Por último, cabe anotar que la estimación de la señal a la salida del filtro acoplado es sesgada, por cuanto

$$\begin{aligned} E \{ \tilde{x}(t) \} &= E \left\{ \int_0^T h_{ac}(\tau; t_0) z(t - \tau) d\tau \right\} = \int_0^T h_{ac}(\tau; t_0) x(t - \tau) d\tau \\ &= x_{1c}(t) \neq x(t) \end{aligned}$$

**Filtros activos y pasivos.** Sea el dispositivo, que realiza la estimación óptima por el criterio del máximo valor de relación señal/ruido de la forma

$$\tilde{x}(t) = \int_T h_{ac}(\tau) z(t - \tau) d\tau \quad (6.65)$$

siendo  $h_{ac}(\tau)$  la solución no homogénea de la ecuación integral (6.65) y  $z(t)$  la combinación aditiva de la señal útil y ruido, observadas en el intervalo  $(t - T, t)$ . La expresión (6.65) se puede interpretar de dos maneras. En el primer caso, el filtro acoplado es un sistema lineal con parámetros constantes, cuya respuesta a impulso  $h_{ac}(\tau)$  se determina por la forma de la señal y de la función de correlación del ruido, acordes con la descripción (6.61). En el caso particular del ruido blanco Gaussiano, la respuesta a impulso se obtiene de la representación espejo de la señal útil.

Sea un generador, que trabaja por el principio descrito por la solución (6.61), entonces, la estimación  $\tilde{x}(t)$  se obtiene mediante el dispositivo de correlación, en el cual la función  $h_{ac}(\tau)$ , obtenida de un generador local, se multiplica la observación de entrada y retenida, mientras el producto obtenido se integra por todo el intervalo de observación. Esta clase de dispositivos para la extracción de señales se denominan *filtros activos*, a diferencia de un sistema lineal con respuesta a impulso  $h_{ac}(\tau)$ , que se denomina *filtro pasivo*.

**Ejemplo 6.4.** Sea la señal pulso cuadrado

$$x(t) = \begin{cases} a, & -\tau_p/2 \leq t \leq \tau_p/2 \\ 0, & |t| > \tau_p/2 \end{cases}$$

El espectro del pulso corresponde al calculado en el ejemplo 1.2,  $X(\omega) = a\tau_p \text{sinc}(\omega\tau_p/2)$ . En concordancia con (6.63) se obtiene que

$$H_{ac}(\omega; t_0) = k(t_0) a\tau_p |\text{sinc}(\omega\tau_p/2)| / S_n(\omega)$$

De la anterior expresión se observa que la respuesta de frecuencia del filtro acoplado se define sobre un rango infinito de frecuencias, mientras su desfase difiere del la señal original. En el sentido estricto, tal función de transferencia es imposible de realizar, por lo que en la práctica el acople del filtro se limita hasta el lóbulo principal, en el cual se considera se concentra el proceso de la mayor parte de la energía de la señal, mientras su desfase se hace tanto como se pueda igual a 0.

### Problemas

**Problema 6.5.** Sea la señal pulso cuadrado

$$x(t) = \begin{cases} a/2 (e^{j\omega_0 t} - e^{-j\omega_0 t}), & -\tau_p/2 \leq t \leq \tau_p/2 \\ 0, & |t| > \tau_p/2 \end{cases}$$

Hallar la función de transferencia y respuesta impulso del respectivo filtro acoplado.

**Problema 6.6.** Sea la señal pulso cuadrado modulado

$$x(t) = \begin{cases} a/2 (e^{j\omega_0 t} - e^{-j\omega_0 t}), & -\tau_p/2 \leq t \leq \tau_p/2 \\ 0, & |t| > \tau_p/2 \end{cases}$$

Hallar la función de transferencia y respuesta impulso del respectivo filtro acoplado.

**Problema 6.7.** Sea la señal de banda ancha

$$x(t) = \begin{cases} a/2 \cos(\omega_0 t + (d_i - 1)\pi/2), & 0 \leq t \leq T \\ 0, & 0 > t > T \end{cases}$$

Asumiendo que la sucesión de valores  $\{d_i = -1, 1 \in T_d : i = 1, \dots, N\}$ ,  $T = NT_d$ , corresponde a una serie de valores aleatorios, hallar la función de transferencia y respuesta impulso del respectivo filtro acoplado.

**Problema 6.8.** Sea el filtro acoplado, que trabaja por el principio (6.65). Hallar la respuesta a impulso del filtro acoplado, asumiendo que la DEP de la señal útil es

$$S_x(\omega) = \frac{S_0}{1 + (\omega T)^2}$$

mientras la perturbación es RBG.

# Capítulo 7

## Análisis de sistemas estocásticos

El principal instrumento de análisis de los sistemas dinámicos complejos corresponde a los métodos de modelado matemático. En el diseño de sistemas dinámicos, aunque sea conocida la estructura de aleatoriedad del proceso a la entrada, es importante determinar las propiedades de estimación del sistema. Por lo tanto, se hace necesario ajustar el análisis con métodos de la matemática estadística.

### 7.1. Modelos de sistemas dinámicos con entradas aleatorias

Como se analiza en el numeral §1.4, la descripción del funcionamiento de un sistema dinámico mediante el análisis en espacio de estados (diagrama de fase) emplea el método de ecuaciones diferenciales,

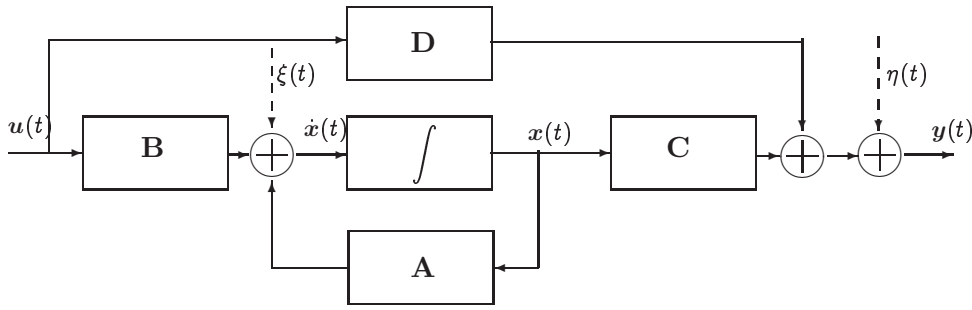
$$\begin{cases} \dot{\mathbf{x}}(t) = \mathbf{f}_x \{ \mathbf{x}(t), \mathbf{u}(t), \boldsymbol{\xi}(t), t \} \\ \mathbf{y}(t) = \mathbf{f}_y \{ \mathbf{x}(t), \mathbf{u}(t), \boldsymbol{\eta}(t), t \} \end{cases} \quad (7.1)$$

donde  $\mathbf{x}_{1 \times n_x}$  es el vector de estados del sistema,  $\mathbf{u}_{1 \times n_u}$  es el vector de la señal de control,  $\mathbf{y}_{1 \times n_y}$  es el respectivo vector de las señales observadas,  $\boldsymbol{\xi}_{1 \times n_\xi}$  es el vector de perturbación en el sistema y  $\boldsymbol{\eta}_{1 \times n_\eta}$  corresponde al vector de errores de medición. Por último,  $t$  es el parámetro escalar de desarrollo del sistema de ecuaciones diferenciales, que típicamente corresponde al tiempo, y el cual toma valores dentro de un intervalo dado,  $t \in T$ .

#### 7.1.1. Sistemas lineales estacionarios

Al asumir la linealidad del sistema en análisis, cuyo diagrama general se muestra en la Figura 7.1, la respectiva descripción (7.1) se simplifica hasta,

$$\begin{cases} \dot{\mathbf{x}}(t) = \mathbf{A}(t) \mathbf{x}(t) + \mathbf{B}(t) \mathbf{u}(t) + \mathbf{I}_\xi(t) \boldsymbol{\xi}(t), \\ \mathbf{y}(t) = \mathbf{C}(t) \mathbf{x}(t) + \mathbf{D}(t) \mathbf{u}(t) + \mathbf{I}_\eta(t) \boldsymbol{\eta}(t), \end{cases} \quad (7.2)$$



**Figura 7.1.** Diagrama general de un sistema dinámico lineal

siendo  $\mathbf{A}(t)$ ,  $\mathbf{B}(t)$ ,  $\mathbf{C}(t)$ ,  $\mathbf{D}(t)$ ,  $\mathbf{I}_\xi(t)$  y  $\mathbf{I}_\eta(t)$  las respectivas matrices de coeficientes de linealidad, que en general, dependen del parámetro  $t$  de desarrollo del sistema. En cuanto a las señales aleatorias,  $\xi(t)$  y  $\eta(t)$ , en los sistemas reales aparecen dificultades relacionadas con la descripción de su influencia en las ecuaciones (7.2).

En el caso simplificado de estudio de un sistema lineal estacionario, la descripción (7.2) se restringe al análisis de las matrices  $\mathbf{A}$ ,  $\mathbf{B}$  y  $\mathbf{C}$  que no dependen del tiempo, luego, para un vector dado de observaciones de salida  $\mathbf{y}$ , el modelo se resume en

$$\begin{cases} \dot{\mathbf{x}} = \mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{u} \\ \mathbf{y} = \mathbf{C}^\top \mathbf{x} \end{cases}$$

con condiciones iniciales  $\mathbf{A}\mathbf{x}(0)$ .

En el numeral §1.4 se analizan tres particularidades de los sistemas lineales, en su descripción mediante el método de ecuaciones diferenciales:

1. Es necesario determinar las condiciones iniciales y las variables de estado,
2. La solución de la ecuación diferencial, que describe el sistema, puede obtenerse mediante un procesador analógico equivalente, el cual se considera el modelo,
3. El modelo del sistema también brinda la posibilidad de realizar el modelado de las señales de salida del sistema.

### 7.1.2. Variables de estado y modelado de procesos aleatorios

La ecuación diferencial, que describe un sistema dinámico lineal, puede ser empleada en el modelado de procesos aleatorios. En el ejemplo 5.1, se muestra el empleo del método de ecuaciones diferenciales en cálculo de los momentos de un circuito  $RC$ . El método de ecuaciones diferenciales ofrece mayores posibilidades para el modelado de procesos aleatorios, como se analiza en el numeral §1.4, la ecuación (1.59) de orden  $m$  puede ser representada en forma alterna por el sistema de ecuaciones de primer orden (1.61) o su variante de descripción vectorial (1.62), a la cual le corresponde la señal de salida (1.64). Del ejemplo 5.1, se puede inferir que este modelo describe procesos, tanto estacionarios como no estacionarios.

Así por ejemplo, en el modelo de la Figura 1.12, sean considerados los correspondientes valores  $t_0 \rightarrow -\infty$ ,  $d^i \eta / dt^i \big|_{t=t_0} = 0$ . Si en calidad de entrada se toma una señal aleatoria vector  $\xi(t)$ , entonces a la salida del sistema se tendrá también otra señal aleatoria  $\eta(t)$ . Ambas componentes vectoriales de las correspondientes señales aleatorias se describen por el sistema (1.70) en la forma,

$$\frac{d\zeta}{dt} = \mathbf{A}(t)\zeta(t) + \mathbf{B}(t)\xi(t), \quad \eta(t) = \mathbf{C}(t)\zeta(t), \quad -\infty < t < \infty$$

En general, si la señal de entrada es ruido blanco Gaussiano, la linealidad del sistema implica un proceso aleatorio de salida también Gaussiano, pero no estacionario. Cabe anotar, que entre más compleja sea la representación de la densidad espectral de potencia que tenga el proceso (mayor grado de aleatoriedad), mayor es la dimensión requerida para el respectivo vector de estado.

**Modelado estocástico.** Se basa en la ecuación diferencial estocástica de la forma,

$$d\mathbf{x} = \mathbf{f}(\mathbf{x}, t) dt + \mathbf{g}(\mathbf{x}, t) d\boldsymbol{\nu} \quad (7.3)$$

donde  $\boldsymbol{\nu}(t)$ ,  $t \in T$ , corresponde a un proceso de Wiener de dimensión  $n$  con función de covarianza de incrementos  $\mathbf{I} dt$ .

Un *proceso escalar de Wiener* se puede definir como un proceso aleatorio  $\nu(t)$ ,  $t \geq t_0$ , con las siguientes propiedades:

1.  $\nu(t_0) = 0$ ,
2.  $\nu(t) \forall t \geq t_0$ , tiene distribución Gaussiana con valor medio igual a cero,
3. El proceso  $\nu(t)$  presenta incrementos estacionarios e independientes, en el sentido en que para cualquier valor de  $t_i \geq t_0$  ( $i = 1, \dots, k$ ), tales que  $t_1 < t_2 < \dots < t_k$ , los correspondientes valores aleatorios:  $\nu(t_1)$ ,  $\nu(t_2) - \nu(t_1)$ ,  $\dots$ ,  $\nu(t_k) - \nu(t_{k-1})$  son mutuamente independientes, mientras la distribución de los incrementos descritos por  $\nu(t) - \nu(s)$ , depende estrictamente de la diferencia de argumentos,  $t - s$ .

De forma similar, se define el proceso de Wiener para el caso de dimensión múltiple, que como consecuencia de las anteriores propiedades, tiene una varianza que crece linealmente en el tiempo, en la forma  $\sigma_\nu^2 = kt$ ,  $k = \text{const.}$ , mientras su función de covarianza se determina mediante la expresión:  $K_\nu(t, s) = k \min(t, s)$ . De (7.3), se tiene que  $\mathbf{x} = \mathbf{x}(t)$  es la solución de la ecuación integral estocástica:

$$\mathbf{x}(t) = \mathbf{x}(t_0) + \int_{t_0}^t \mathbf{f}(\mathbf{x}(\tau), \tau) d\tau + \sum_{t_0}^t \mathbf{g}(\mathbf{x}(\tau), \tau) d\boldsymbol{\nu}(\tau) \quad (7.4)$$

Básicamente, la dificultad en el cálculo de la solución de (7.3) está en la correcta determinación del último término de (7.4). En este sentido, se emplea la integral de Ito, (5.36),

expresada en la forma [11]:

$$\int_{t_0}^t g(\mathbf{x}(\tau), \tau) d\boldsymbol{\nu}(\tau) = \lim_{\Delta t_i \rightarrow 0} \sum_i^N g(\mathbf{x}(t_i), t_i) (\boldsymbol{\nu}(t_{i+1}) - \boldsymbol{\nu}(t_i))$$

siendo  $\Delta t_i = t_{i+1} - t_i$ . La media y covarianza de la solución, respectivamente, son:

$$E\{\mathbf{x}(t)\} = E\{\mathbf{x}(t_0)\} + E\left\{\int_{t_0}^t f(\mathbf{x}(s), s) ds\right\}$$

$$\text{cov}\{\mathbf{x}(t + \varepsilon) - \mathbf{x}(t) | \mathbf{x}(t)\} = g(\mathbf{x}, t) g^T(\mathbf{x}, t) \varepsilon + o(\varepsilon)$$

donde  $o(\varepsilon)$  es un valor de orden infinitamente pequeño con relación a  $\varepsilon$ , tal que  $o(\varepsilon)/\varepsilon \rightarrow 0$  cuando  $\varepsilon \rightarrow 0$ .

Sea la función  $\mathbf{y}(\mathbf{x}, t)$ , que se supone continuamente diferenciable por  $t$  y dos veces continuamente diferenciable por  $\mathbf{x}$ , donde  $\mathbf{x}$  cumple la ecuación (7.3). Entonces,  $\mathbf{y}(\mathbf{x}, t)$  cumple la siguiente ecuación diferencial:

$$\begin{aligned} d\mathbf{y} &= \frac{\partial \mathbf{y}}{\partial t} dt + \sum_{i=1}^n \frac{\partial \mathbf{y}}{\partial x_i} + \frac{1}{2} \sum_{i,j,k=1}^n \frac{\partial^2 \mathbf{y}}{\partial x_i \partial x_j} g_{ik} g_{jk} dt \\ &= \left( \frac{\partial \mathbf{y}}{\partial t} + \sum_{i \neq 1}^n \frac{\partial \mathbf{y}}{\partial x_i} f_i + \frac{1}{2} \sum_{i,j,k=1}^n \frac{\partial^2 \mathbf{y}}{\partial x_i \partial x_j} g_{ik} g_{jk} \right) dt + \sum_{i=1}^n \frac{\partial \mathbf{y}}{\partial x_i} (gd\boldsymbol{\nu})_i \end{aligned} \quad (7.5)$$

siendo  $g_{ij}$  los elementos de la matriz  $\mathbf{g}(\mathbf{x}, t)$  y  $(gd\boldsymbol{\nu})_i$  es el elemento  $i$  del vector  $(gd\boldsymbol{\nu})$ . La ecuación (7.5) presenta ya una forma implementable para la solución de la ecuación diferencial estocástica. Así por ejemplo, de manera particular se analiza para (7.3) la siguiente ecuación lineal diferencial, dada en forma vectorial,

$$d\mathbf{x} = \mathbf{A}(t) \mathbf{x} dt + d\boldsymbol{\nu} \quad (7.6)$$

donde  $\mathbf{x} = \mathbf{x}(t)$  es un vector con dimensión  $n$ ,  $\mathbf{A}(t)$  es una matriz cuadrada de orden  $n \times n$ , mientras  $\boldsymbol{\nu}(t), t \in T$  es un proceso de Wiener de dimensión  $n$  y función de covarianza  $K_{\boldsymbol{\nu}} dt$ . Además, al suponer que el valor inicial  $\mathbf{x}(t_0)$  es un valor aleatorio Gaussiano con valor medio  $\mathbf{m}_{1\mathbf{x}_0}$  y función de covarianza  $K_{\mathbf{x}_0}$ , entonces, la ecuación diferencial estocástica (7.6) se escribe en la forma integral equivalente como

$$\mathbf{x}(t) = \boldsymbol{\Phi}(t, t_0) \mathbf{x}(t_0) + \int_{t_0}^t \boldsymbol{\Phi}(t, s) d\boldsymbol{\nu}(s)$$

La matriz  $\boldsymbol{\Phi}$  cumple la ecuación diferencial (con condiciones iniciales  $\boldsymbol{\Phi}(t_0, t_0) = \mathbf{I}$ ):

$$\frac{d}{dt} \boldsymbol{\Phi}(t, t_0) = \mathbf{A}(t) \boldsymbol{\Phi}(t, t_0)$$



La solución de la ecuación estocástica diferencial (7.6) corresponde a un proceso aleatorio con valor medio  $E\{\mathbf{x}(t)\} = \mathbf{m}_{1\mathbf{x}}(t)$  y función de covarianza  $K_{\mathbf{x}}(s, t)$ , donde

$$\begin{aligned} dm_{1\mathbf{x}}/dt &= \mathbf{A}(t) m_{1\mathbf{x}}(t_0) \quad m_{1\mathbf{x}}(t_0) = m_{1\mathbf{x}_0} \\ K_{\mathbf{x}}(s, t) &= \begin{cases} \Phi(s, t) \mathbf{P}(t), & s \geq t \\ \mathbf{P}(s) \Phi(t, s), & s < t \end{cases} \end{aligned}$$

En cuanto a la matriz  $\mathbf{P}$ , ésta cumple la siguiente ecuación diferencial,

$$\frac{d\mathbf{P}}{dt} = \mathbf{A}\mathbf{P} + \mathbf{P}\mathbf{A}^T + \mathbf{K}_{\nu}, \quad \mathbf{P}(t_0) = \mathbf{K}_{\mathbf{x}_0}$$

De otra parte, cuando el conjunto de valores sobre los cuales se determina el intervalo de  $t$  corresponde a los enteros, esto es, se tiene la malla de valores  $T \in \{\dots, -1, 0, 1, \dots\}$ , el sistema determinístico, ya discreto en el tiempo, se describe por la ecuación de iteraciones,

$$\begin{cases} \mathbf{x}[t+1] &= \mathbf{f}_x\{\mathbf{x}[t], \mathbf{u}[t], t\} \\ \mathbf{y}[t] &= \mathbf{f}_y\{\mathbf{x}[t], \mathbf{u}[t], t\} \end{cases}$$

donde  $\mathbf{x}[t]$  es el vector de estados de dimensión  $n$ ,  $\mathbf{y}[t]$  es el vector de observaciones de dimensión  $l$ , mientras  $\mathbf{u}[t]$  es el vector de control. Al agregar factores aditivos de perturbación, el modelo anterior se completa hasta la forma,

$$\begin{cases} \mathbf{x}[t+1] &= \mathbf{f}_x\{\mathbf{x}[t], \mathbf{u}[t], t\} + \mathbf{f}_{\xi}\{\mathbf{x}[t], \xi[t], t\} \\ \mathbf{y}[t] &= \mathbf{f}_y\{\mathbf{x}[t], \mathbf{u}[t], t\} + \mathbf{f}_{\eta}\{\mathbf{x}[t], \eta[t], t\} \end{cases}$$

donde  $\mathbf{f}_{\xi}\{\cdot\}$  y  $\mathbf{f}_{\eta}\{\cdot\}$  son los funcionales para las perturbaciones del modelo. La simplificación en la solución del modelo, se obtiene en forma práctica, asumiendo la dependencia lineal entre la función, en general, vectorial  $\mathbf{f}_x$  y la señal  $\mathbf{x}[t]$ , además asumiendo que el control del sistema se realiza al escoger adecuadamente los elementos de  $\mathbf{f}_x$ , por lo que la siguiente ecuación estocástica diferencial lineal describe el vector de estados del sistema:

$$\mathbf{x}[t+1] = \Phi[t+1, t] \mathbf{x}[t] + \mathbf{e}[t] \quad (7.7)$$

Sean los vectores  $\mathbf{e}[t]$  y  $\mathbf{e}[s]$  Gaussianos e independientes,  $\forall t \neq s$ , con los momentos:

$$E\{\mathbf{e}[t]\} = 0, \quad E\{\mathbf{e}[t] \mathbf{e}^T[s]\} = \mathbf{R}_{\mathbf{e}}(t, s)$$

Así mismo, se considera el estado inicial del sistema  $\mathbf{x}[t_0]$  una variable aleatoria Gaussiana con valor medio  $\mathbf{m}_{1\mathbf{x}_0}$  y función de covarianza  $\mathbf{K}_{\mathbf{x}_0}$ . Entonces, la solución de la ecuación lineal estocástica (Gaussiana) de iteraciones (7.7) corresponde a un proceso también Gaussiano con valor medio que cumple la relación (siendo  $\Phi$  alguna matriz de orden  $n \times n$ ):

$$\mathbf{m}_{1\mathbf{x}}[t+1] = \Phi[t+1, t] \mathbf{m}_{1\mathbf{x}}[t]$$

con condición inicial  $\mathbf{m}_{1\mathbf{x}}[t_0]$  y función de covarianza

$$K_{\mathbf{x}}(s, t) = \Phi[s, t] \mathbf{P}[t], \quad \forall s \leq t$$

donde la matriz  $\mathbf{P}(t)$  se halla de la relación iterativa

$$\mathbf{P}[t+1] = \Phi[t+1, t] \mathbf{P}[t] \Phi^T[t+1, t] + \mathbf{K}_{\mathbf{x}}$$

con condiciones iniciales  $\mathbf{P}[t] = \mathbf{K}_{\mathbf{x}_0}$ . Si las matrices  $\Phi$  y  $\mathbf{K}_{\mathbf{x}}$  son constantes en el tiempo, entonces, a partir de todas las anteriores condiciones derivadas del modelo (7.7), se tiene,

$$\mathbf{P}[t] = \Phi^T \mathbf{K}_{\mathbf{x}_0} (\Phi^T)^t + \sum_{s=0}^{t-1} \Phi^s \mathbf{K}_{\mathbf{x}} (\Phi^T)^s$$

**Modelado discreto de sistemas estocásticos continuos.** En este caso, se tiene el cambio de las ecuaciones estocásticas diferenciales a las respectivas ecuaciones de iteración.

Sean los procesos, de funcionamiento del sistema  $\mathbf{x}(t)$  y de obtención de las observaciones  $\mathbf{y}(t)$ , descritos por las ecuaciones estocásticas diferenciales vectoriales de la forma,

$$\begin{cases} d\mathbf{x} = \mathbf{A}\mathbf{x}dt + d\mathbf{v}_1 \\ d\mathbf{y} = \mathbf{C}\mathbf{x}dt + d\mathbf{v}_2 \end{cases} \quad (7.8)$$

donde  $\mathbf{x}_{1 \times n}$  es el vector de estados del sistema, mientras  $\mathbf{y}_{1 \times r}$  es el vector de todas las observaciones,  $\{\mathbf{v}_m(t) : m = 1, 2\}$ , son los vectores los procesos de Wiener con dimensiones respectivas  $n$  y  $r$ , y funciones de covarianza de incremento:  $\mathbf{K}_{\mathbf{v}_m}dt$ . Se asume, además, que las variables de salida se observan a lo largo de la malla de valores discretos de tiempo  $\{t_k : k = 0, 1, \dots\}$ . Entonces, los valores de las variables de estado y de las variables de salida observadas para las ecuaciones diferenciales estocásticas (7.8), sobre una malla discreta de tiempo  $\{t_k : k = 0, 1, \dots\}$ , están relacionadas por las ecuaciones de diferencias estocásticas:

$$\begin{cases} \mathbf{x}(t_{i+1}) = \Phi \mathbf{x}(t_i) + \hat{\mathbf{v}}_1(t_i) \\ \mathbf{z}(t_{i+1}) = \mathbf{y}(t_{i+1}) - \mathbf{y}(t_i) = \mathbf{S}\mathbf{x}(t_i) + \hat{\mathbf{v}}_2(t_i) \end{cases}$$

donde la matriz  $\Phi = \Phi(t_{i+1}, t_i), \forall i \geq 0$ , se determina por las relaciones

$$\frac{d}{dt} \Phi(t, t_i) = \mathbf{A}(t) \Phi(t, t_i), \quad \Phi(t_i, t_i) = \mathbf{I}, \quad t_i \leq t \leq t_{i+1},$$

mientras, la matriz  $\mathbf{S} = \mathbf{S}(t_{i+1}, t_i), \forall i \geq 0$ , se da por la expresión,

$$\mathbf{S}(t_{i+1}, t_i) = \int_{t_i}^{t_{i+1}} \mathbf{C}(s) \Phi(s, t_i) ds \quad (7.9)$$

Los vectores  $\{\hat{\mathbf{v}}_m(t_i) : m = 1, 2; \forall i \geq 0\}$  corresponden a sucesiones de valores aleatorios

Gaussianos independientes con valor medio igual a cero y con las correspondientes funciones de correlación, determinadas a partir de las siguientes expresiones ( $\forall i \geq 0$ ):

$$\mathbf{R}_{\boldsymbol{\nu}_1}(t_i) = \int_{t_i}^{t_{i+1}} \boldsymbol{\Phi}(t_{i+1}, s) \mathbf{R}_{\boldsymbol{\nu}_1} \boldsymbol{\Phi}^T(t_{i+1}, s) ds \quad (7.10a)$$

$$\mathbf{R}_{\boldsymbol{\nu}_2}(t_i) = \int_{t_i}^{t_{i+1}} \left( \mathbf{S}(t_{i+1}, s) \mathbf{R}_{\boldsymbol{\nu}_1} \mathbf{S}^T(t_{i+1}, s) + \mathbf{R}_{\boldsymbol{\nu}_2}(s) \right) ds \quad (7.10b)$$

Finalmente, la función de covarianza mutua entre los procesos  $\{\boldsymbol{\nu}_m : m = 1, 2\}$  se determina por la expresión,

$$\mathbf{K}_{\boldsymbol{\nu}_1 \boldsymbol{\nu}_2}(t_i) = E \left\{ \boldsymbol{\nu}_1(t_i) \boldsymbol{\nu}_2^T(t_i) \right\} = \int_{t_i}^{t_{i+1}} \boldsymbol{\Phi}(t_i, s) \mathbf{R}_{\boldsymbol{\nu}_1}(s) \mathbf{S}^T(t_{i+1}, s) ds, \quad i \geq 0$$

Las ecuaciones iterativas estocásticas (7.9), (7.10a), (7.10b), mas las ecuaciones diferenciales estocásticas (7.8), desde el punto de vista de sus propiedades estadísticas dentro del intervalo de discretización, son idénticas, con lo cual el respectivo proceso de funcionamiento continuo se puede simular en los sistemas de proceso digital.

### 7.1.3. Estimación de parámetros en sistemas estocásticos

La necesidad de aumentar la efectividad del funcionamiento de los sistemas de control, en condiciones de factores aleatorios, estimuló el desarrollo de métodos de optimización orientada a mejorar el registro y precisión de la información sobre las propiedades de la planta a diseñar, así como de los regímenes de su funcionamiento.

El trabajo de un sistema se debe representar por algún modelo, por ejemplo, los descritos en el numeral §7.1, en el cual, el vector de estados registra las medidas de los diferentes dispositivos de captura de información sobre el estado del sistema, dicho de otra manera, es el vector de mediciones.

Sea el objeto de análisis un sistema dinámico en tiempo discreto, que se describe por las siguientes ecuaciones de diferencia:

$$\mathbf{x}[k+1] = \mathbf{A}\mathbf{x}[k], \quad k \geq 0, \quad k \in \mathbb{Z} \quad (7.11)$$

siendo  $\mathbf{x}$  un vector con dimensión  $n$  y  $\mathbf{A}$  una matriz cuadrada de orden  $n \times n$ . Se asume que el proceso de medición se describe por

$$\mathbf{y}[k] = \mathbf{C}\mathbf{x}[k] \quad (7.12)$$

donde  $\mathbf{y}$  es el vector de medidas con dimensión  $r$ , mientras la matriz  $\mathbf{C}$  tiene orden  $r \times n$ .

La planta se considera *observable* si a partir de los registros de medida  $\mathbf{y}[0], \dots, \mathbf{y}[n-1]$  se puede determinar el estado  $\mathbf{x}[0]$  y, por lo tanto, todos los demás subsecuentes estados

del sistema,  $\mathbf{x}[1], \mathbf{x}[2], \dots$

Al describir de forma secuencial la ecuación (7.12), para los valores de  $k = 0, \dots, n-1$  se obtiene la siguiente expresión matricial,

$$\begin{bmatrix} \mathbf{y}^T[0] : \mathbf{y}^T[1] : \dots : \mathbf{y}^T[n-1] \end{bmatrix} = \mathbf{x}^T[0] \mathbf{H} \quad (7.13)$$

donde  $\mathbf{H} = [\mathbf{C}^T : \mathbf{A}^T \mathbf{C}^T : \dots : \mathbf{A}^{T[n-1]} \mathbf{C}^T]$ .

En general, la condición de observabilidad se puede hacer equivalente a la condición de existencia y unicidad de la solución  $\mathbf{x}[0]$  en (7.13), que consiste en que el rango de la matriz  $\mathbf{H}$  es igual a  $n$ . Entonces, la pareja de matrices  $\mathbf{A}, \mathbf{C}$  se denomina *observable*.

El concepto de *identificación*, que está asociado con la determinación de la matriz  $\mathbf{A}$  (dados los valores medidos del vector de estados), es equivalente a la condición de existencia y unicidad de la solución  $\mathbf{A}$  de la ecuación matricial,

$$\begin{bmatrix} \mathbf{x}[1] : \dots : \mathbf{x}[n] \end{bmatrix} = \mathbf{A} \begin{bmatrix} \mathbf{x}[0] : \dots : \mathbf{A}^{n-1} \mathbf{x}[0] \end{bmatrix} = \mathbf{A} \mathbf{S}$$

lo que implica que la matriz  $\mathbf{S}$  tenga rango  $n$ .

Las anteriores condiciones pueden ser generalizadas para el caso de un sistema dinámico en tiempo continuo, para todo valor  $t_0 \leq t \leq T$ , cuyo funcionamiento y proceso de medida se dan por las respectivas ecuaciones diferenciales,

$$\begin{cases} \dot{\mathbf{x}}(t) = \mathbf{A}(t) \mathbf{x}(t) \\ \mathbf{y}(t) = \mathbf{C}(t) \mathbf{x}(t) \end{cases}$$

Si se nota la matriz de transición del sistema lineal  $\dot{\mathbf{x}}(t) = \mathbf{A}(t) \mathbf{x}(t)$  por  $\Phi(\tau, t)$ , entonces la condición de observabilidad consiste en el cumplimiento de la siguiente relación:

$$\mathbf{M}(t, t_0) = \int_{t_0}^t \Phi^T(\tau, t) \mathbf{C}^T(\tau) \mathbf{C}(\tau) \Phi(\tau, t) d\tau > 0$$

La *matriz de observabilidad*  $\mathbf{M}(t, t_0)$ , mediante la cual se puede determinar el vector  $\dot{\mathbf{x}}(t)$ :  $\mathbf{x}(t) = \mathbf{M}^{-1}(t, t_0) \mathbf{y}(t)$ , se halla de la ecuación diferencial,

$$\frac{d}{dt} \mathbf{M}(t, t_0) = -\mathbf{A}^T(t) \mathbf{M}(t, t_0) - \mathbf{M}(t, t_0) \mathbf{A}(t) + \mathbf{C}^T(t) \mathbf{C}(t)$$

con condición inicial  $\mathbf{M}(t, t_0) = \mathbf{0}$ .

De forma similar, la anterior descripción de la condición de observabilidad se mantiene para los sistemas dinámicos en tiempo discreto. Sea  $\Phi[i, j]$  la matriz de transición del sistema (7.11), tal que,

$$\mathbf{x}[k] = \Phi[k, 0] \mathbf{x}[0], \quad \Phi[0, 0] = \mathbf{I}$$

Si se formula la matriz de observabilidad en la forma,

$$\mathbf{M}[k, 0] = \sum_{l=0}^k \Phi^T[l, k] \mathbf{C}^T[l] \mathbf{C}[l] \Phi[l, k]$$

entonces, la condición de observabilidad se escribe en forma de la desigualdad:  $\mathbf{M}[k, 0] > 0$ .

En la práctica, durante el proceso de registro de señales ocurren errores aleatorios de medida, así mismo, el funcionamiento del propio sistema dinámico se ve afectado por perturbaciones, también de naturaleza aleatoria. Además, la cantidad de registros de medida a procesar pueden ser insuficientes, o por el contrario, estar en número excesivo. Por lo tanto, el principal problema consiste en la obtención de las estimaciones óptimas para las variables que describen, tanto los parámetros, como el estado del sistema dinámico. Asociada a la estimación óptima de los valores de medida está la filtración, que se analiza en el capítulo §6. Sin embargo, también relacionada con la tarea de filtración está el problema de la determinación de los parámetros del sistema o *identificación*. Aunque son suficientemente conocidas las diversas técnicas de filtración óptima de sistemas dinámicos, su efectividad está condicionada por la correspondencia que existe entre los parámetros del modelo y la planta real de análisis, que no siempre es suficientemente precisa. En este sentido, se han mejorado las técnicas clásicas de filtración para el caso de datos a priori dados de forma no adecuada, que nutren los algoritmos de filtración. El mejoramiento de los algoritmos se puede alcanzar con el empleo de técnicas estables de estimación, basadas en algoritmos que se adaptan a las condiciones de cambio externas.

#### 7.1.4. Identificación de parámetros en sistemas lineales

Uno de los métodos más efectivos, al mismo tiempo más cómodos de representación en sistemas de proceso digital, es el *método de Kaczmarz* para la identificación de parámetros en sistemas de dimensión múltiple. En particular, sea el sistema lineal discreto en la forma,

$$\mathbf{y}[k] = \mathbf{c}^T \mathbf{x}[k], \quad k = 1, 2, \dots \quad (7.14)$$

donde  $\mathbf{y}[k]$  es el escalar de la señal de salida del sistema,  $\mathbf{x}[k]$  es el vector de la señal de entrada y  $\mathbf{c}$  es un vector de parámetros desconocidos del sistema ( $\mathbf{c}$  y  $\mathbf{x}$  con dimensión  $m$ ), de tal manera, que se pueden estimar los valores del vector de parámetros  $\mathbf{c}$  del sistema a partir de las sucesiones de observaciones  $\{\mathbf{x}[k] : k > 1\}$  y  $\{\mathbf{y}[k] : k > 1\}$ . La efectividad del método principalmente depende de las cualidades estocásticas de la sucesión de entrada  $\{\mathbf{x}[k]\}$ . Así por ejemplo, bajo ciertas presunciones y asumiendo la ortogonalidad mutua de los vectores  $\mathbf{x}[1], \dots, \mathbf{x}[m]$ , entonces el valor del vector  $\mathbf{c}$  puede ser obtenido con suficiente precisión en solo  $m$  pasos de cálculo. Además, si los mismos vectores, siendo procesos aleatorios, son Gaussianos con igual varianza, entonces la estimación del vector  $\mathbf{c}$  demanda una muestra de volumen  $l = (4 \dots 5)$  términos.

El método de Kaczmarz tiene amplio uso en la solución de problemas de identificación de sistemas lineales no estacionarios y su algoritmo de cálculo tiene estructura simple, basada

en las siguientes relaciones recurrentes:

$$\tilde{\mathbf{c}}[k] = \tilde{\mathbf{c}}[k-1] + \frac{\mathbf{y}[k] - \tilde{\mathbf{c}}^T[k-1] \mathbf{x}[k]}{\mathbf{x}^T[k] \mathbf{x}[k]} \mathbf{x}[k], \quad k = 2, 3, \dots \quad (7.15)$$

donde  $\tilde{\mathbf{c}}[k]$  es la estimación del vector de parámetros desconocidos en el paso  $k$ . El valor inicial del vector  $\tilde{\mathbf{c}}[1]$  se fija de manera a priori.

La interpretación geométrica de la relación (7.15) consiste en que cada estimación  $\tilde{\mathbf{c}}[k]$  puede considerarse como la proyección del valor previamente calculado de  $\tilde{\mathbf{c}}[k-1]$  sobre el respectivo  $k$  hiperplano en un espacio euclídeo  $\mathbb{R}^m$ , determinado por la relación (7.14). En este caso, la sucesión de las normas de los vectores  $\Delta \mathbf{c}[k] = \tilde{\mathbf{c}}[k] - \mathbf{c}[k]$  converge, decreciendo de forma monótona, hasta cero. Sin embargo, la presunción de independencia estadística, implica que en la práctica, al aumentar la correlación entre valores cercanos en el tiempo de la sucesión de entrada  $\mathbf{x}[1], \dots, \mathbf{x}[m]$ , de manera significativa disminuye la efectividad del algoritmo de Kaczmarz. En este sentido, se han propuesto varias modificaciones al algoritmo básico (7.15), que además permiten aumentar su velocidad de proceso [45, 46].

El algoritmo generalizado de estimación recurrente de Kaczmarz se puede describir por las siguientes expresiones recurrentes:

$$\begin{aligned} \tilde{\mathbf{c}}[k] &= \tilde{\mathbf{c}}[k-1] + \frac{\mathbf{y}[k] - \tilde{\mathbf{c}}^T[k-1] \mathbf{x}[k]}{\mathbf{x}^T[k] \mathbf{x}[k]} \mathbf{x}[k], \quad k = 2, 3, \dots \\ \hat{\mathbf{c}}[k] &= \tilde{\mathbf{c}}[k] + \alpha[k] (\hat{\mathbf{c}}[k-2] - \hat{\mathbf{c}}[k-1]) \end{aligned}$$

siendo  $\hat{\mathbf{c}}[k]$  la estimación actual del vector de parámetros  $\mathbf{c}$  y  $\alpha$  el factor de corrección, que se calcula como,

$$\alpha[k] = \begin{cases} 1, & \rho^2 \{\hat{\mathbf{c}}[k-2], \hat{\mathbf{c}}[k]\} > (\rho^2 \{\hat{\mathbf{c}}[k-2], \hat{\mathbf{c}}[k-1]\} - \mu)^2 \\ 0, & \text{caso contrario} \end{cases}$$

donde

$$\rho\{\hat{\mathbf{c}}[m], \hat{\mathbf{c}}[n]\} = \frac{\mathbf{y}[n] - \mathbf{c}^T[m] \mathbf{x}[n]}{\mathbf{x}^T[n] \mathbf{x}[n]}$$

El factor  $\mu$  es igual a

$$\mu = \mathbf{y}[k] - \hat{\mathbf{c}}^T[k-1] \mathbf{x}[k] / \mathbf{x}^T[k] \mathbf{x}[k]$$

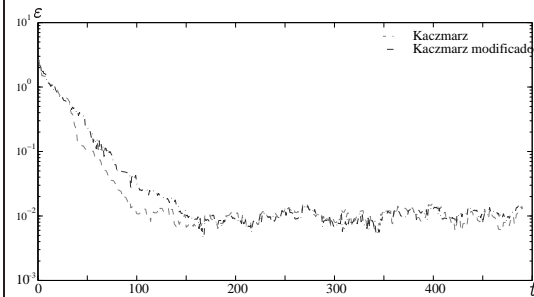
para el caso del algoritmo [46] y  $\mu = 0$  para el algoritmo presentado en [45].

De esta manera, el algoritmo de Kaczmarz consiste en el procedimiento recurrente con los valores iniciales desconocidos  $\hat{\mathbf{c}}[1]$ ,  $\hat{\mathbf{c}}[2]$  y  $\hat{\mathbf{c}}[2]$ , los cuales deben ser fijados a priori.

**Ejercicio en el CP 7.1.** La identificación de los coeficientes de un filtro no recursivo de orden  $N$ , se puede realizar mediante el algoritmo de Kaczmarz (en sus versiones simple o modificada), generando los respectivos vectores de señal entrada y salida del filtro, como se muestra a continuación:

```
function c = Kaczmarz(x,y,N)
% Algoritmo de Kaczmarz
% Entrada x, salida y, orden N
c = ones(N);
for i=N:length(x),
    ye = c'*x(i-N+1:i);
    R = x(i-N+1:i)'*x(i-N+1:i);
    c = c + ( y(i) - ye ) * R^-1 * x(i-N+1:i);
end;

function c = KaczmarzMod(x,y,N)
% Algoritmo de Kaczmarz modificado
c = ones(N); c1 = ones(N); c2 = ones(N);
for i=N+1:L,
    c2 = c1;    c1 = c;
    R = x(i-N+1:i)'*x(i-N+1:i);
    mu = (y(i) - we_1'*x(i-N+1:i))*R^-1;
    rho_1 = (y(i) - we_2'*x(i-N+1:i))*R^-1;
    rho_2 = (y(i-1) - we_2'*x(i-N:i-1))
    x0 = (x(i-N:i-1)'*x(i-N:i-1))^-1;
    rho_2 = rho_2*x0;
    if rho_1^2 > (rho_2^2-mu)^2,
        alpha = 1;
    else
        alpha = 0;
    end;
    c0 = ( y(i) - c1'*x(i-N+1:i) );
    c = c1 + c0*R^-1*x(i-N+1:i);
    c = c + alpha*( c2 - c1 );
end;
```



La convergencia de ambos procedimientos del algoritmo de Kaczmarz, para el ejemplo de identificación de un filtro de orden  $N = 10$ , se muestra en la Figura de la parte derecha.

## Problemas

**Problema 7.1.** Mediante el análisis de los polos de las respectivas funciones de transferencia  $H(z)$ , comprobar para qué valores de  $k \in [0, 1]$  se cumple la condición de estabilidad de cada uno de los siguientes filtros:

$$H(z) = \frac{1 - z^{-1}}{1 - kz^{-1}}$$

$$H(z) = \frac{1 - z^{-2}}{1 - kz^{-1}}$$

$$H(z) = \frac{50 - 76.5z^{-1} + 10.575z^{-2} + 26.581z^{-3} - 5.882z^{-4}}{1 + 0.8z^{-1} - 0.9z^{-2} - kz^{-3} + 0.9144z^{-4} + 0.5184z^{-5}}$$

$$H(z) = \frac{1 - kz^{-1}}{1 - z^{-1}}$$

$$H(z) = \frac{1 - z^{-2}}{1 - 1.8z^{-1} + kz^{-2}}$$

**Problema 7.2.** Estimar los coeficientes de la función respuesta a impulso de los filtros pasabajos, descritos en los problemas 1.17 y 1.18, empleando los algoritmos de Kaczmarz (simple y modificado). Hallar el error absoluto de aproximación.





## 7.2. Filtración de sistemas dinámicos

### 7.2.1. Filtración Wiener

Sea la entrada de un sistema, el proceso vectorial aleatorio,  $\mathbf{x}(t)$ , de dimensión  $n$  y con valor medio 0, el cual se debe restablecer de la manera más precisa posible, a partir de la estimación  $\hat{\mathbf{x}}(t)$ , que se asume puede ser descrita por el siguiente sistema de ecuaciones diferenciales con coeficientes variables en el tiempo:

$$\frac{d\hat{\mathbf{x}}}{dt} = \mathbf{P}(t)\hat{\mathbf{x}}(t) + \mathbf{Q}(t)\mathbf{z}(t) \quad (7.16)$$

donde  $\mathbf{z}(t)$  es la entrada de dimensión  $n$  del sistema, (7.16), la cual corresponde a un proceso aleatorio del tipo no estacionario con valor medio igual a 0, el vector  $\hat{\mathbf{x}}(t)$  es de dimensión  $n$ , mientras las matrices  $\mathbf{P}(t)$  y  $\mathbf{Q}(t)$  son de orden  $n \times n$  y  $n \times m$ , respectivamente.

El error de representación se determina por la diferencia simple entre la señal deseada y la observación para su estimación,

$$\mathbf{e}(t) = \mathbf{x}(t) - \hat{\mathbf{x}}(t) \quad (7.17)$$

Sin embargo, la calidad del filtro (7.16) se estima por el criterio,

$$J(t) = E \left\{ \mathbf{e}^T(t) \mathbf{e}(t) \right\} \quad (7.18)$$

que se minimiza para optimizar el desempeño del sistema. La parte derecha de (7.18) es la suma de los elementos diagonales de la respectiva matriz,  $\text{trace}\{E\{\mathbf{e}^T(t) \mathbf{e}(t)\}\}$ .

De otra parte, teniendo en cuenta (7.16), la señal  $\mathbf{x}(t)$  se determina mediante la relación,

$$\hat{\mathbf{x}}(t) = \int_{t_0}^t \boldsymbol{\mu}(t, \tau) \mathbf{z}(\tau) d\tau$$

siendo  $\boldsymbol{\mu}(t, \tau)$  una matriz de orden  $n \times m$ , determinada por el producto,

$$\boldsymbol{\mu}(t, \tau) = \boldsymbol{\Phi}(t, \tau) \mathbf{Q}(\tau) \quad (7.19)$$

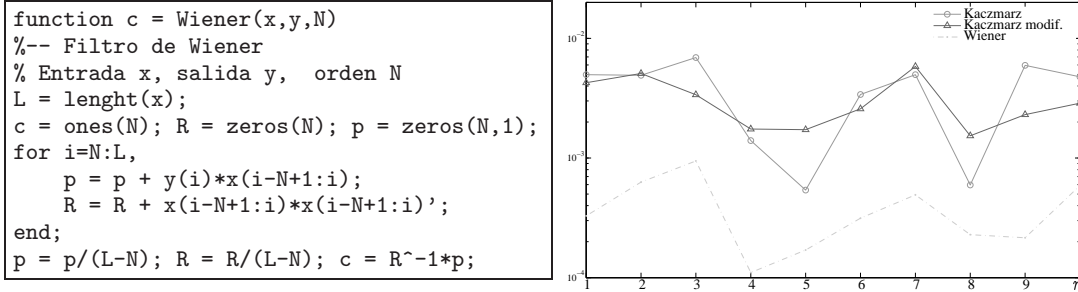
donde  $\boldsymbol{\Phi}(t, \tau) = \mathbf{f}(t) \mathbf{f}^{-1}(\tau)$ ,  $\mathbf{f}$  es la matriz fundamental de soluciones del sistema de ecuaciones diferenciales homogéneas en la forma,  $d\hat{\mathbf{x}}/dt = \mathbf{P}(t)\hat{\mathbf{x}}(t)$ , con condiciones iniciales 0. La selección óptima de la matriz  $\boldsymbol{\mu}(t, \tau)$  de (7.19), que corresponde a la función de peso del filtro (7.16), se minimiza por el criterio  $J(t)$  en (7.18).

Cabe anotar, que la condición de optimalidad de  $J(t)$  es equivalente a la ecuación integral matricial de Wiener-Hopf [47]:

$$E \left\{ \mathbf{x}(t) \mathbf{z}^T(t_1) \right\} = \int_{t_0}^t \boldsymbol{\mu}(t, t_2) E \left\{ \mathbf{z}(t_2) \mathbf{z}^T(t_1) \right\} dt_2, \quad t_0 \leq t_1 \leq t \quad (7.20)$$

En la práctica, la solución de la ecuación integral (7.20), además de exigir el conocimiento a priori de la estructura estocástica de las señales (funciones de correlación y espectros de potencia), demanda grandes recursos computacionales.

**Ejercicio en el CP 7.2.** La identificación del sistema, en el ejemplo 7.1, se realiza mediante el filtro de Wiener, empleando el procedimiento seguidamente ilustrado. Los resultados en el error de aproximación obtenidos en los casos de los algoritmos de Kaczmarz y de Wiener se muestran en la Figura de la parte derecha para valor de  $N = 10$ .



### 7.2.2. Filtración óptima de sistemas discretos lineales

La estimación del vector de estados del sistema sobre el conjunto de observaciones a la salida del sistema, teniendo en cuenta las perturbaciones presentes y los errores de medida, conlleva a una tarea de filtración, que en estos casos, uno de los métodos más efectivos corresponde a la *filtración de Kalman* [48].

Sea el modelo de un sistema dinámico lineal en tiempo discreto,

$$\mathbf{x}[k+1] = \mathbf{A}[k]\mathbf{x}[k] + \mathbf{w}[k], \quad k = 0, 1, \dots \quad (7.21)$$

donde  $\mathbf{x}[k]$  es el vector de estados del sistema que no puede ser medido de forma directa,  $\mathbf{w}[k]$  son las perturbaciones aleatorias con estructura Gaussiana, siendo  $\mathbf{x}$  y  $\mathbf{w}$  de dimensión  $n$ . La matriz determinística  $\mathbf{A}[k]$  es de orden  $n \times n$ .

Las siguientes características del modelo se consideran fijadas a priori:

$$E\{\mathbf{x}[0]\} = \mathbf{x}[0], \quad E\{(\mathbf{x}[0] - \bar{\mathbf{x}}[0])(\mathbf{x}[0] - \bar{\mathbf{x}}[0])^T\} = \mathbf{P}[0]$$

$$E\{\mathbf{w}[k]\} = \mathbf{w}[k], \quad E\{(\mathbf{w}[k] - \bar{\mathbf{w}}[k])(\mathbf{w}[k] - \bar{\mathbf{w}}[k])^T\} = \mathbf{Q}[k]\delta_{kl}$$

$$E\{(\mathbf{w}[k] - \bar{\mathbf{w}}[k])(\mathbf{x}[0] - \bar{\mathbf{x}}[0])^T\} = 0$$

siendo  $\mathbf{Q}[k]$  una matriz conocida.

Además, se considera que la relación, que describe el proceso de medida del vector de

estados, es también lineal,

$$\mathbf{y}[k] = \mathbf{C}[k] \mathbf{x}[k] + \mathbf{v}[k], \quad k = 0, 1, \dots$$

donde  $\mathbf{y}[k]$  es el vector de medida de las respectivas observaciones,  $\mathbf{v}[k]$  es el vector aleatorio Gaussiano de error de medida, ambos vectores  $\mathbf{v}$  e  $\mathbf{y}$  tienen dimensión  $r$  y  $\mathbf{C}[k]$  es una matriz dada de orden  $r \times n$ . Del vector  $\mathbf{v}[k]$ , se considera conocida su función de correlación propia  $\mathbf{R}_v[k]$ , además, se asumen sus siguientes momentos:

$$\begin{aligned} E\{\mathbf{v}[k]\} &= \bar{\mathbf{v}}[k], \quad E\{\mathbf{v}[k] \mathbf{v}^T[k]\} = \mathbf{R}_v[k] \delta_{kl} \\ E\{(\mathbf{w}[k] - \bar{\mathbf{w}}[k]) \mathbf{v}^T[k]\} &= 0, \quad E\{(\mathbf{x}[k] - \bar{\mathbf{x}}[k]) \mathbf{v}^T[k]\} = 0, \end{aligned}$$

Los valores determinados anteriormente, implican que el vector  $\mathbf{v}[k]$  tiene correlación nula, tanto con las perturbaciones del sistema, como con su vector inicial de condiciones. A efectos de simplificar el modelo, sin pérdidas de generalidad, se puede considerar que  $\mathbf{v}[k] = \mathbf{w}[k] = 0$ , mientras las condiciones iniciales se pueden asumir  $\mathbf{x}[0] = 0$ .

Basados en la sucesión de observaciones  $\mathbf{y}[0], \dots, \mathbf{y}[k]$  es necesario hallar la estimación insesgada  $\hat{\mathbf{x}}[k]$ , cuya error  $\mathbf{e}[k]$  se da por la expresión (7.17) cuando el criterio de calidad de la estimación es el valor cuadrático medio,  $E\{\mathbf{e}^2[k]\}$ . En calidad de estimación óptima del vector  $\hat{\mathbf{x}}[k]$  se toma su esperanza matemática condicional [48]:

$$\hat{\mathbf{x}}[k] = E\{\mathbf{x}[k] | \mathbf{y}[0], \dots, \mathbf{y}[k]\}$$

con la cual se obtiene el mínimo valor cuadrático medio del error de estimación.

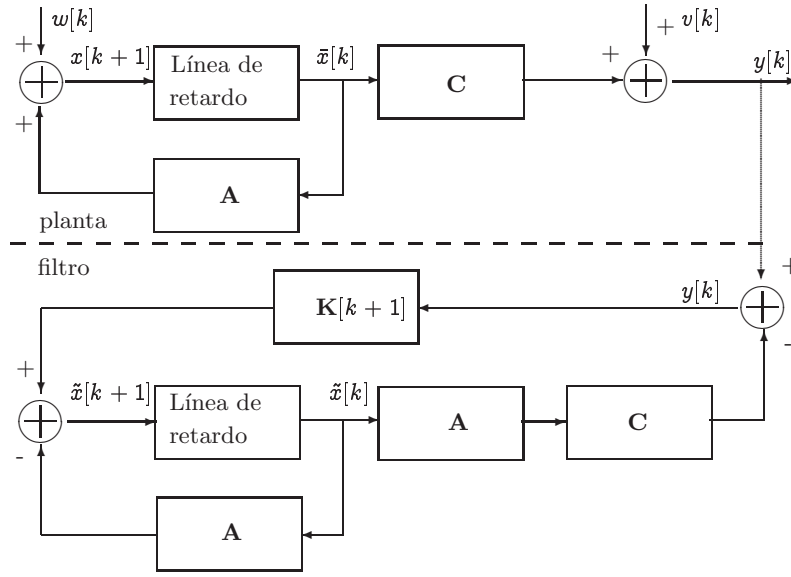
El algoritmo de Kalman, que se describe de forma recurrente, determina el proceso de evolución en el tiempo de la estimación óptima  $\hat{\mathbf{x}}[k]$ :

$$\begin{cases} \hat{\mathbf{x}}[k] &= \mathbf{A}[k-1] \hat{\mathbf{x}}[k-1] \\ \tilde{\mathbf{x}}[k] &= \hat{\mathbf{x}}[k] + \mathbf{K}[k] (\mathbf{y}[k] - \mathbf{C}[k] \hat{\mathbf{x}}[k]), \quad k = 1, 2, \dots \end{cases} \quad (7.22)$$

donde  $\hat{\mathbf{x}}[0]$  se asume conocido. Las demás expresiones recurrentes necesaria en (7.22) se determinan como:

$$\begin{cases} \bar{\mathbf{P}}[k] &= \mathbf{A}[k-1] \hat{\mathbf{P}}[k-1] \mathbf{A}^T[k-1] + \mathbf{Q}[k-1], \\ \mathbf{K}[k] &= \bar{\mathbf{P}}[k] \mathbf{C}^T[k] (\mathbf{C}[k] \bar{\mathbf{P}}[k] \mathbf{C}^T[k] + \mathbf{R}[k])^{-1}, \\ \hat{\mathbf{P}}[k] &= \bar{\mathbf{P}}[k] - \mathbf{K}[k] \mathbf{C}[k] \bar{\mathbf{P}}[k] = (\mathbf{I} - \mathbf{K}[k] \mathbf{C}[k]) \bar{\mathbf{P}}[k], \quad \bar{\mathbf{P}}[0] = \mathbf{P}[0] \end{cases}$$

El diagrama funcional del filtro de Kalman se ilustra en la Figura 7.2, en la cual se muestra que el procedimiento consiste en que la estimación de  $\mathbf{x}[k]$  corresponde a la estimación extrapolada del estado del sistema después de  $k-1$  pasos de medida. La estimación  $\tilde{\mathbf{x}}[k]$  es la corrección del valor estimado  $\hat{\mathbf{x}}[k]$ , obtenido como resultado del proceso de la  $k$  medida. A su vez, las matrices  $\mathbf{P}[k]$  y  $\bar{\mathbf{P}}[k]$  corresponden, respectivamente a la matriz extrapolada y a la corregida, de las funciones de covarianza de los errores de filtración. Por



**Figura 7.2.** Diagrama funcional del filtro Kalman.

cierto, la matriz de covarianza de los errores de estimación  $\tilde{\mathbf{P}}[k]$  no depende de la sucesión de medidas  $\mathbf{y}[0], \dots, \mathbf{y}[k]$ , por lo tanto, sus valores pueden ser calculados de antemano y llevados a alguna posición de memoria.

El modelo del sistema dinámico de control se puede hacer un tanto más complejo, en la siguiente forma,

$$\mathbf{x}[k+1] = \mathbf{A}[k]\mathbf{x}[k] + \mathbf{B}[k]\mathbf{u}[k] + \mathbf{w}[k]$$

En este caso, la sucesión  $\mathbf{y}[0], \dots, \mathbf{y}[k]$  se determina a partir de la expresión

$$\mathbf{y}[k] = \mathbf{C}[k]\mathbf{x}[k] + \mathbf{v}[k]$$

En cuanto a la sucesión  $\mathbf{u}[0], \dots, \mathbf{u}[k]$ , ésta corresponde a la sucesión determinística dada de control.

De esta manera, el filtro de Kalman se determina por las siguientes relaciones recurrentes:

$$\tilde{\mathbf{x}}[k] = \mathbf{A}[k-1]\tilde{\mathbf{x}}[k-1] + \mathbf{B}[k-1]\mathbf{u}[k-1]$$

$$\hat{\mathbf{x}}[k] = \tilde{\mathbf{x}}[k] + \mathbf{K}[k](\mathbf{y}[k] - \mathbf{C}[k]\tilde{\mathbf{x}}[k])$$

El vector  $\mathbf{x}[0]$  es conocido de antemano y la matriz  $\mathbf{K}$  se calcula, tal cual se hace para el procedimiento descrito anteriormente.

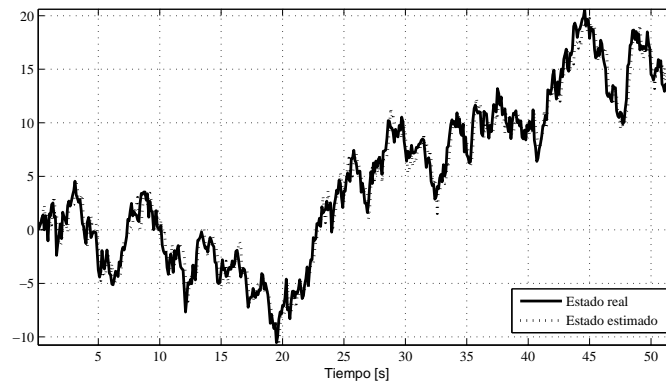
**Ejercicio en el CP 7.3.** Desarrollar el algoritmo de estimación mediante el filtro Kalman para un proceso que describe una caminata aleatoria en la forma:

$$\begin{aligned}x[k+1] &= x[k] + w[k], \\ y[k] &= x[k] + v[k]\end{aligned}$$

siendo las señales aleatorias  $w$  y  $v$ , RBG con momentos:

$$\begin{aligned}E\{w\} &= E\{v\} = 0 \\ \sigma_v &= \sigma_w = 1\end{aligned}$$

```
% filtro de Kalman en un sistema discreto
%--- Inicialización de parámetros de simulación
A = 1;           % Matriz de transición de estado A
C = 1;           % Matriz de medición de estado C
R = 1;           % Matriz de covarianza del ruido de medición
Q = 1;           % Matriz de covarianza de la perturbación del estado
tf = 52;         % Tiempo final de simulación
h = 0.1;         % Tiempo de muestreo
L = tf/h;
%--- Generación de la señal medida
x(1) = 0;        % Estado inicial del sistema
y(1) = x(1) + R*randn(1); % Medición inicial
for i=1:L-1,
    x(i+1) = A*x(i) + Q*randn(1);
    y(i+1) = C*x(i) + R*randn(1);
end;
%--- Inicio de estimación de Kalman
xe(1) = 0;       % Estimación inicial del estado del sistema
P = 1;           % Matriz de covarianza inicial del error
for i=2:L,
    xm = A*xe(i-1);
    Pm = A*P*A' + Q;
    Pxy = Pm*C';
    Pyy = C*Pm*C' + R;
    GK = Pxy*Pyy^-1;
    ym = C*xm;
    xe(i) = xm + GK*( y(i) - ym );
    P = Pm - GK*C*Pm;
end;
```



La anterior Figura muestra una trayectoria estimada del filtro discreto de Kalman, mediante el algoritmo presentado.

### 7.2.3. Filtración óptima de sistemas continuos lineales

Además del cambio usual de las ecuaciones de diferencias a ecuaciones diferenciales, o de sumatorias a integrales, el modelo de filtración continuo, como generalización del discreto, es más complejo, en la medida en que la cantidad de observaciones en unidad de tiempo, en el caso continuo no es finita, sino infinita. En calidad de modelo de estudio, se tiene las siguientes ecuaciones de la planta y el dispositivo de medida (filtro de *Kalman-Bucy*):

$$\begin{cases} \dot{\mathbf{x}}(t) = \mathbf{A}(t)\mathbf{x}(t) + \mathbf{w}(t) \\ \mathbf{y}(t) = \mathbf{C}(t)\mathbf{x}(t) + \mathbf{v}(t) \end{cases}$$

para  $\forall t \geq t_0$ . El vector a estimar  $\mathbf{x}(t)$  es de dimensión  $n$ , para el cual se asume que  $\mathbf{x}(t_0)$  es un vector aleatorio con media cero y función de covarianza,

$$\mathbf{R}_{\mathbf{x}_0}(t_0) = E \left\{ (\mathbf{x}(t_0) - E\{\mathbf{x}(t_0)\}) (\mathbf{x}(t_0) - E\{\mathbf{x}(t_0)\})^\top \right\}$$

mientras, el vector de medida  $\mathbf{y}(t)$  tiene dimensión  $r$ . Las matrices  $\mathbf{A}$  y  $\mathbf{C}$  son conocidas. Para las perturbaciones  $\mathbf{w}(t)$ , con dimensión  $n$ , y los errores Gaussianos de medida, con dimensión  $r$ , se cumplen los siguientes momentos:

$$\begin{aligned} E\{\mathbf{w}(t)\} &= 0, E\{\mathbf{w}(t)\mathbf{w}(t)^\top\} = \mathbf{R}_w(t)\delta(t-\tau), \\ E\{\mathbf{v}(t)\} &= 0, E\{\mathbf{v}(t)\mathbf{v}(t)^\top\} = \mathbf{R}_v(t)\delta(t-\tau) \end{aligned}$$

Las matrices  $\mathbf{R}_w$  y  $\mathbf{R}_v$  son simétricas y con elementos variables en el tiempo. Las señales aleatorias  $\mathbf{w}(t)$  y  $\mathbf{v}(t)$  tienen función de correlación mutua cero. La estimación  $\hat{\mathbf{x}}(t)$  se debe hallar a partir de los resultados de medida  $\mathbf{y}(\tau)$ ,  $t_0 \leq \tau \leq t$ , que en el caso de la estimación lineal no sesgada, con el mínimo error cuadrático medio, se describe por la ecuación,

$$\frac{d}{dt}\hat{\mathbf{x}}(t) = \mathbf{A}(t)\hat{\mathbf{x}}(t) + \mathbf{K}(t)(\mathbf{y}(t) - \mathbf{C}(t)\hat{\mathbf{x}}(t)), \quad \hat{\mathbf{x}}(t_0) = 0,$$

La matriz de Kalman se calcula de la siguiente expresión,

$$\mathbf{K}(t) = \hat{\mathbf{R}}_{\mathbf{x}_0}(t) \mathbf{C}^\top(t) \mathbf{R}_v^{-1}(t) \quad (7.23)$$

La matriz de covarianza del error se determina por la *ecuación matricial diferencial de Riccati*

$$\begin{aligned} \frac{d}{dt}\hat{\mathbf{R}}_{\mathbf{x}_0}(t) &= \mathbf{A}(t)\hat{\mathbf{R}}_{\mathbf{x}_0}(t) + \hat{\mathbf{R}}_{\mathbf{x}_0}(t)\mathbf{A}^\top(t) - \hat{\mathbf{R}}_{\mathbf{x}_0}(t)\mathbf{C}^\top(t)\mathbf{R}_v^{-1}(t)\mathbf{C}(t)\hat{\mathbf{R}}_{\mathbf{x}_0}(t) + \mathbf{R}_w(t), \end{aligned} \quad (7.24)$$

asumiendo  $\hat{\mathbf{R}}_{\mathbf{x}_0}(t_0) = \mathbf{R}_{\mathbf{x}_0}(t_0)$ .

Si el modelo matemático está dado en la forma

$$\begin{cases} \dot{\mathbf{x}}(t) + \mathbf{A}(t)\mathbf{x}(t) + \mathbf{B}(t)\mathbf{u}(t) + \mathbf{w}(t), & E\{\mathbf{x}(t_0)\} = \xi \\ \mathbf{y}(t) = \mathbf{C}(t)\mathbf{x}(t) + \mathbf{v}(t) \end{cases}$$

entonces la estimación lineal no sesgada del vector de estados  $\mathbf{x}(t)$ , que cumple la condición del mínimo error cuadrático medio, se determina como la solución de la siguiente ecuación diferencial,

$$\frac{d}{dt}\tilde{\mathbf{x}}(t) = \mathbf{A}(t)\tilde{\mathbf{x}}(t) + \mathbf{B}(t)\mathbf{u}(t) + \mathbf{K}(t)(\mathbf{y}(t) - \mathbf{C}(t)\tilde{\mathbf{x}}(t)),$$

asumiendo

$$\mathbf{x}(t_0) = \xi$$

La matriz  $\mathbf{K}(t)$  se determina por la expresión (7.23), pero la matriz  $\mathbf{R}_{\mathbf{x}_0}(t)$  corresponde a la solución de la ecuación de Riccati en (7.24), la cual analíticamente tiene solución en muy pocos casos, por lo tanto, se considera su solución por métodos numéricos [49].

Uno de los casos, cuando la ecuación tiene solución, corresponde al siguiente. Sea la ecuación diferencial homogénea del error de filtración de la forma,

$$\dot{\tilde{\mathbf{x}}}(t) = (\mathbf{A}(t) - \hat{\mathbf{R}}_{\mathbf{x}_0}(t)\mathbf{C}^\top(t)\mathbf{R}_v^{-1}(t)\mathbf{C}(t))\tilde{\mathbf{x}}(t)$$

con su conjugado (si  $\dot{\tilde{\mathbf{x}}}(t) = \mathbf{F}(t)\tilde{\mathbf{x}}(t)$ , entonces la matriz conjugada será  $\dot{\xi}(t) = -\mathbf{F}^\top(t)\xi(t)$ ), de lo cual se obtiene,

$$\dot{\xi}(t) = (-\mathbf{A}^\top(t) + \mathbf{C}^\top(t)\mathbf{R}_v^{-1}(t)\mathbf{C}(t)\hat{\mathbf{R}}_{\mathbf{x}_0}(t))\xi(t)$$

La ecuación (7.24) se multiplica por  $\xi(t)$ , por lo que se obtiene,

$$\begin{aligned} & \frac{d}{dt}\hat{\mathbf{R}}_{\mathbf{x}_0}(t)\xi(t) \\ &= (\mathbf{A}(t)\hat{\mathbf{R}}_{\mathbf{x}_0}(t) + \hat{\mathbf{R}}_{\mathbf{x}_0}(t)\mathbf{A}^\top(t) - \hat{\mathbf{R}}_{\mathbf{x}_0}(t)\mathbf{C}^\top(t)\mathbf{R}_v^{-1}(t)\mathbf{C}(t)\hat{\mathbf{R}}_{\mathbf{x}_0}(t) + \mathbf{R}_w(t))\xi(t) \end{aligned}$$

luego, multiplicando la ecuación conjugada por  $\hat{\mathbf{R}}_{\mathbf{x}_0}(t)$  y reemplazando en la anterior ecuación, se tiene

$$\hat{\mathbf{R}}_{\mathbf{x}_0}(t)\dot{\xi}(t) + \hat{\mathbf{R}}_{\mathbf{x}_0}(t)\xi(t) = (\mathbf{A}(t)\hat{\mathbf{R}}_{\mathbf{x}_0}(t) + \mathbf{R}_w(t)\xi(t))$$

Como se observa, la última ecuación no presenta el término no lineal, entonces, al introducir la siguiente notación se obtiene un sistema lineal homogéneo del tipo,

$$\frac{d}{dt}\begin{bmatrix} \xi \\ \eta \end{bmatrix} = \begin{bmatrix} -\mathbf{A}^\top(t) & \mathbf{C}^\top(t)\mathbf{R}_v^{-1}(t)\mathbf{C}(t) \\ \mathbf{R}_w(t) & \mathbf{A}(t) \end{bmatrix}$$

donde los valores de  $\xi(t_0)$  y  $\eta(t_0)$  están dados a través de  $\mathbf{R}_{\mathbf{x}_0}(t)$ . Si se hallan los términos  $\xi(t)$  y  $\eta(t)$ , entonces al existir  $\xi^{-1}(t)$  se obtiene

$$\hat{\mathbf{R}}_{\mathbf{x}_0}(t) = \eta(t) \xi^{-1}(t)$$

En general, los conceptos de observabilidad y controlabilidad se pueden complementar de forma mutua, gracias al principio de dualidad de Kalman, según el cual la observación y la filtración del sistema inicial,

$$\begin{cases} \dot{\mathbf{x}}(t) &= \mathbf{A}(t) \mathbf{x}(t) + \mathbf{B}(t) \mathbf{u}(t), \\ \mathbf{y}(t) &= \mathbf{C}(t) \mathbf{x}(t), \quad t_0 \leq t \end{cases}$$

corresponden al control y regulación de un sistema dual, y viceversa.

El modelo matemático del sistema dual está dado por las expresiones

$$\begin{cases} -\dot{\xi}(t) &= \mathbf{A}^\top(t) \xi(t) + \mathbf{C}^\top(t) \mathbf{v}(t) \\ \eta(t) &= \mathbf{B}^\top(t) \xi(t) \end{cases}$$

de las cuales se puede demostrar que las propiedades del control óptimo permanecen vigentes, por lo menos desde el punto de vista algorítmico, en la síntesis de filtros óptimos [11].

En general, la tarea conjunta de estimación de los parámetros de la señal y de los estados del sistema es de carácter no lineal. Así sea el modelo analizado descrito por las relaciones,

$$\begin{cases} \dot{\mathbf{x}} &= \mathbf{f}_x \{\mathbf{x}, \mathbf{u}, \mathbf{a}, \xi, t\}; \\ \mathbf{y} &= \mathbf{f}_y \{\mathbf{x}, \mathbf{u}, \mathbf{C}, \eta, t\} \end{cases}$$

donde  $\mathbf{x}(t)$  es el vector de estados,  $\mathbf{u}(t)$  es la señal observada de entrada,  $\mathbf{a}(t)$  es el vector de parámetros de la planta,  $\xi(t)$  es el ruido de entrada o perturbación,  $\eta(t)$  es el error de medida,  $\mathbf{y}(t)$  es la señal observada de salida y  $\mathbf{C}$  es la matriz de observaciones.

El vector de parámetros de la planta  $\mathbf{a}$  se interpreta como la función dada por el control  $\dot{\mathbf{a}} = 0$  para un vector de condiciones iniciales  $\mathbf{a}(0)$ , luego, el modelo es,

$$\begin{bmatrix} \dot{\mathbf{x}} \\ \dot{\mathbf{a}} \end{bmatrix} = \begin{bmatrix} \mathbf{f} \{\mathbf{x}, \mathbf{u}, \mathbf{a}, \xi, t\} \\ 0 \end{bmatrix}$$

que se puede transformar en la forma

$$\begin{bmatrix} \dot{\mathbf{x}} \\ \dot{\mathbf{a}} \end{bmatrix} = \mathbf{M} \begin{bmatrix} \mathbf{x} \\ \mathbf{a} \end{bmatrix}$$

donde  $\mathbf{M}$  no depende ni de  $\mathbf{x}$  ni de  $\mathbf{a}$ . Entonces, inclusive para una planta lineal con dependencia lineal de parámetros, la tarea de estimación conjunta de los parámetros de la señal y de los estados del sistema tiene carácter no lineal con relación al vector de parámetros y al estado del sistema  $[\mathbf{x}^\top, \mathbf{a}^\top]$ .



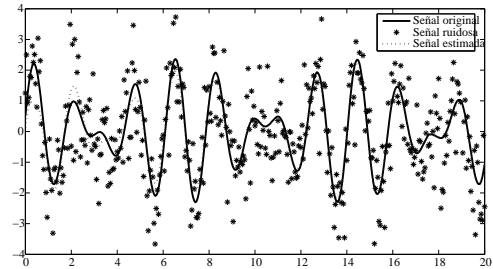
**Ejercicio en el CP 7.4.** Realiza un estimador, empleando la Filtración Kalman-Bucy, para un oscilador cuasiperiódico descrito por las siguientes ecuaciones diferenciales:

$$\begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \\ \dot{x}_3 \\ \dot{x}_4 \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ -\omega_1^2 & -2a & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & -\omega_2^2 & -2a \end{bmatrix} \begin{bmatrix} x_1(t) \\ x_2(t) \\ x_3(t) \\ x_4(t) \end{bmatrix} + \begin{bmatrix} 0 \\ 1 \\ 0 \\ 1 \end{bmatrix} w(t)$$

$$y(t) = \begin{bmatrix} 1 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} x_1(t) \\ x_2(t) \\ x_3(t) \\ x_4(t) \end{bmatrix} + v(t)$$

donde  $a = 0$ ,  $\omega_1 = 4$  y  $\omega_2 = \pi$ .

```
function KalmanBucy
% Oscilador armónico amortiguado
% Condiciones iniciales
xo = [0.5 4 0 4]'; xeo = zeros(4,1);
Po = 20*eye(4); r = 1;
ko=[xo;xeo;Po(:,1);Po(:,2);Po(:,3);Po(:,4)];
h = 5e-2; tspan = 0:h:20;
[t,k] = ode45(@KB,tspan,ko);
y = k(:,1)+k(:,3);
yn = y + r*randn(length(tspan),1);
ye = k(:,5)+k(:,7);
%-----
function dk = KB(t,k)
%-- Ajustes generales del filtro de Kalman
w1 = 4; w2 = pi; a = 0; r = 1; q = 0.1;
%-- Matrices del sistema
A = [ 0 1 0 0;
      -w1^2 -2*a 0 0;
      0 0 0 1;
      0 0 -w2^2 -2*a];
C = [1 0 1 0];
%-- Matriz de covarianza
Q = q*eye(4); Q(1,1) = 0; Q(3,3) = 0;
G = [0 1 0 1];
%-- Sistema real
dx = zeros(4,1); x = k(1:4); dx =
A*x + G*q*randn(4,1);
y = C*x + r*randn(1);
%-- Ecuación de Ricatti
P = [k(9:12) k(13:16) k(17:20) k(21:24)];
dP = zeros(4); K = P*C'*r^-1;
dP = A*P + P*A' - K*r*K' + Q;
%-- Estado estimado del sistema real
dxs = zeros(4,1); xs = k(5:8);
dxs = A*xs + K*(y-C*xs);
%-- Salida de datos
dk = zeros(20,1); dk(1:4) = dx;
dk(5:8) = dxs; dk(9:12) = dP(:,1);
dk(13:16) = dP(:,2); dk(17:20) = dP(:,3);
dk(21:24) = dP(:,4);
%--
```



### 7.2.4. Filtración de sistemas no lineales

Sea un sistema lineal en tiempo continuo descrito por la ecuación estocástica vectorial del tipo difusivo,

$$d\mathbf{x}(t) = \mathbf{f}\{\mathbf{x}, t\} dt + \mathbf{v}\{\mathbf{x}, t\} d\beta(t), \quad t \geq t_0 \quad (7.25)$$

en las cuales  $\beta(t)$  es un proceso estándar de Wiener, en el sentido en que los valores aleatorios son reales con distribución normal de incrementos, valor medio  $E\{d\beta\} = 0$  y varianza  $E\{d\beta^2\} = dt$ . Las funciones  $\mathbf{f}\{\cdot\}$  y  $\mathbf{v}\{\cdot\}$  son respectivamente los coeficientes de transporte y difusión.

El vector de estados del sistema  $\mathbf{x}(t)$  no puede ser observado directamente. Solamente es observable el vector  $\mathbf{y}(t)$ , por lo que el proceso de medida se describe también por otra ecuación diferencial estocástica,

$$d\mathbf{y}(t) = \mathbf{h}(\mathbf{x}, t) dt + d\gamma(t) \quad (7.26)$$

Los vectores  $\mathbf{x}$  y  $\mathbf{f}$  tienen dimensión  $n$ , mientras los vectores  $\mathbf{h}$  e  $\mathbf{y}$  son de dimensión  $r$ . Los procesos tipo Wiener  $\beta(t)$  y  $\gamma(t)$  se asumen con correlación mutua nula.

La estimación del vector de estados del sistema se puede realizar mediante la linealización de las ecuaciones (7.25) y (7.26), con el consecutivo empleo de métodos de filtración lineal, lo cual conlleva a las siguientes ecuaciones:

$$\begin{cases} \dot{\tilde{\mathbf{x}}}(t) &= \mathbf{f}\{\tilde{\mathbf{x}}, t\} + \mathbf{v}\{\tilde{\mathbf{x}}, t\} \beta(t) + \mathbf{R}_{\mathbf{x}} \left( \frac{\partial \mathbf{h}}{\partial \mathbf{x}} \right)^T \mathbf{R}_{\gamma}^{-1} (\mathbf{y}(t) - \mathbf{h}(\tilde{\mathbf{x}}, t)), \\ \dot{\mathbf{R}}_{\mathbf{x}} &= \frac{\partial \mathbf{f}}{\partial \mathbf{x}} \mathbf{R}_{\mathbf{x}_0} + \dot{\mathbf{R}}_{\mathbf{x}} \left( \frac{\partial \mathbf{f}}{\partial \mathbf{x}} \right)^T + \mathbf{v} \mathbf{R}_{\beta} \mathbf{v}^T - \mathbf{R}_{\mathbf{x}} \left( \frac{\partial \mathbf{h}}{\partial \mathbf{x}} \right)^T \mathbf{R}_{\gamma}^{-1} \frac{\partial \mathbf{h}}{\partial \mathbf{x}} \mathbf{R}_{\mathbf{x}} \end{cases}$$

con los siguientes momentos

$$\begin{aligned} E\{\mathbf{x}(t_0)\} &= \tilde{\mathbf{x}}(t_0) = \bar{\mathbf{x}}(t_0) \\ E\left\{(\mathbf{x}(t_0) - \bar{\mathbf{x}}(t_0))(\mathbf{x}(t_0) - \bar{\mathbf{x}}(t_0))^T\right\} &= \mathbf{R}_{\mathbf{x}}(t_0) = \mathbf{R}_{\mathbf{x}_0} \\ E\left\{(\dot{\beta}(t) - \bar{\dot{\beta}}(t))(\dot{\beta}(\tau) - \bar{\dot{\beta}}(\tau))^T\right\} &= \mathbf{R}_{\beta}(t) \delta(t - \tau) \\ E\left\{(\dot{\gamma}(t) - \bar{\dot{\gamma}}(t))(\dot{\gamma}(\tau) - \bar{\dot{\gamma}}(\tau))^T\right\} &= \mathbf{R}_{\gamma}(t) \delta(t - \tau) \\ E\left\{(\mathbf{x}(t_0) - \bar{\mathbf{x}}(t_0))(\dot{\beta}(\tau) - \bar{\dot{\beta}}(\tau))^T\right\} &= 0 \end{aligned}$$

siendo  $\bar{\mathbf{x}}(t_0)$ ,  $\bar{\dot{\gamma}}(t)$  y  $\bar{\dot{\beta}}(t)$  son los valores promedios de los respectivos procesos.

En el caso de análisis de la filtración de sistemas no lineales en tiempo discreto, se parte del proceso no lineal de dimensión múltiple, descrito en la forma,

$$\mathbf{x}[k+1] = \mathbf{f}^k\{\mathbf{x}[k]\} + \mathbf{w}[k], \quad k = 0, 1, \dots$$

para el cual, las medidas se realizan mediante un dispositivo, descrito por la ecuación no lineal,  $\mathbf{y}[k] = \mathbf{h}^k \{\mathbf{x}[k]\} + \mathbf{v}[k]$ ,  $k = 0, 1, \dots$

Las presunciones sobre la estructura de aleatoriedad de todas las variables estocásticas siguen siendo las mismas que se hicieron para el filtro Kalman-Bucy. El algoritmo de filtración no lineal se determina de las siguientes relaciones recurrentes:

$$\begin{aligned}\tilde{\mathbf{x}}[k] &= \bar{\mathbf{x}}[k] + \mathbf{K}[k] - \mathbf{h}^k \{\bar{\mathbf{x}}[k]\}, \quad \bar{\mathbf{x}}[k] = \mathbf{f}^k \{\tilde{\mathbf{x}}[k-1]\}, \\ \mathbf{R}_x[k] &= \frac{\partial \mathbf{f}^{k-1}}{\partial \mathbf{x}[k-1]} \hat{\mathbf{R}}[k-1] \left( \frac{\partial \mathbf{f}^{k-1}}{\partial \mathbf{x}[k-1]} \right)^\top + \mathbf{R}_w[k-1] \\ \mathbf{K}[k] &= \hat{\mathbf{R}}_x[k] \left( \frac{\partial \mathbf{h}^k}{\partial \mathbf{x}[k]} \right)^\top \left( \frac{\partial \mathbf{h}^k}{\partial \mathbf{x}[k]} \hat{\mathbf{R}}_x[k] \left( \frac{\partial \mathbf{h}^k}{\partial \mathbf{x}[k]} \right)^\top + \mathbf{R}_v[k] \right)^{-1} \\ \hat{\mathbf{R}}_x[k] &= \bar{\mathbf{R}}_x[k] - \mathbf{K}[k] \frac{\partial \mathbf{h}^k}{\partial \mathbf{x}[k]} \bar{\mathbf{R}}_x[k]\end{aligned}$$

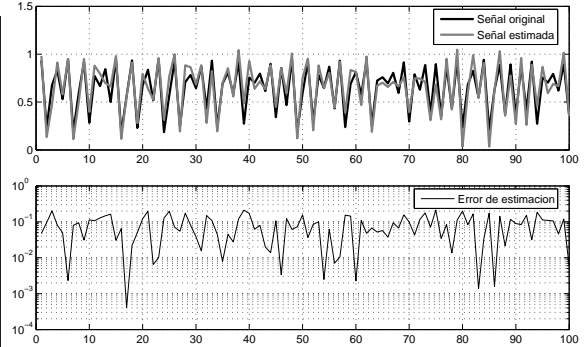
en las que se asume:  $\bar{\mathbf{R}}_x[0] = \mathbf{R}_x[0]$ , además, con condiciones iniciales conocidas,  $\tilde{\mathbf{x}}[0]$ .

El método de linealización es suficientemente efectivo, mientras no se tengan perturbaciones de potencia significativa, las que afectan la calidad de la estimación. Otro factor a tener en cuenta, en el rendimiento del filtro, está en la selección apropiada de los valores iniciales del vector estimación  $\tilde{\mathbf{x}}[0]$ .

**Ejercicio en el CP 7.5.** Realizar el filtro extendido de Kalman para la estimación de estado del mapa logístico, el cual se define mediante la siguiente ecuación de diferencias no lineal:

$$x[k] = \alpha(x[k-1] - x[k-1]^2)$$

```
%--Estimación del mapa logístico
function LMEstEKF
%-- Se genera la serie temporal
alpha = 3.8;
L = 5000;
x = zeros(L,1);
x(1) = 0.11;
for i=2:L,
    x(i) = alpha*x(i-1) - alpha*x(i-1)^2;
end;
y = x + 0.1*randn(L,1);
%-- Filtro extendido de Kalman
xe = zeros(L,1); xe(1) = 1;
P = 1e0;
Rw = 1e-10; Rv = 0.1;
for i=2:L,
    xe_ = alpha*xe(i-1) - alpha*xe(i-1)^2;
    A = alpha - 2*alpha*xe(i-1);
    Px_ = A*P*A' + Rw;
    K = Px_ * ( Px_ + Rv )^-1;
    xe(i) = xe_ + K*( y(i) - xe_ );
    Px = ( 1 - K )*Px_;
end;
%--
```



### 7.2.5. Filtración adaptativa de sistemas dinámicos

La filtración de Kalman, en variadas situaciones prácticas, puede ser no suficientemente efectivo, esto es, no brindar la suficiente precisión, o bien, simplemente no converger. Lo anterior es debido, entre otras, por las siguientes razones:

- Los parámetros en el modelo del sistema no concuerdan suficientemente con valores reales, durante su funcionamiento, por la dificultad en la estimación,
- La presunción de Gaussividad sobre la estructura, tanto en las perturbaciones como en los errores de medida, no es cierta en manera significativa.

Por lo tanto, es necesario mejorar el algoritmo de Kalman, de tal manera que se adapte a los factores a priori dados de no concordancia en el modelo, además a las diversas estructuras de aleatoriedad de las señales aleatorias durante el proceso.

Los métodos actuales para la resolución de esta tarea se orientan en dos aproximaciones básicas:

- La estimación, más bien la identificación, de los parámetros del modelo en el proceso de funcionamiento y el subsecuente empleo de las estimaciones obtenidas en el filtro convencional del Kalman.
- El análisis del carácter de evolución del sistema y los procesos aleatorios de medida y su adaptación por parte del filtro mediante cambios en su estructura de funcionamiento.

La segunda orientación, siendo más compleja en su implementación, se considera más efectiva, y es la que se analiza en este caso. Sea un sistema lineal en tiempo discreto dado en la forma,

$$\begin{cases} \mathbf{x}[k+1] &= \mathbf{A}[k] \mathbf{x}[k] + \mathbf{G}[k] \mathbf{w}[k] \\ \mathbf{y}[k] &= \mathbf{C}[k] \mathbf{x}[k] + \gamma[k] \alpha[k] + (1 - \gamma[k]) \mathbf{p}[k], \end{cases} \quad \mathbf{A}, \mathbf{C}, \mathbf{G} \in \mathbb{R}; \quad k = 0, 1, \dots$$

donde  $\mathbf{A}[k]$ ,  $\mathbf{C}[k]$  y  $\mathbf{G}[k]$  son matrices determinísticas de orden  $n \times n$ ,  $n \times r$  y  $n \times s$ , respectivamente,  $\mathbf{x}[k]$  es el vector de estados del sistema con dimensión  $n$ ,  $\mathbf{w}[k]$  son las perturbaciones aleatorias con dimensión  $s$ , independientes entre si, de estructura Gaussiana con valor medio cero y matriz de covarianza  $\mathbf{K}_{\mathbf{w}[k]}$ ,  $\mathbf{y}[k]$  es el vector de medidas con dimensión  $r$ .

Las magnitudes aleatorias independientes entre si, que forman la sucesión  $\gamma[k]$ , toman los valores 0 y 1 con probabilidades  $g_k$  y  $p_k$ :  $\mathbf{P}\{\gamma[k] = 1\} = p_k = 1 - g_k$ . Los vectores aleatorios  $\alpha[k]$  y  $\beta[k]$ , ambos de dimensión  $r$ , son Gaussianos con valor medio cero y función de covarianza  $\mathbf{K}_{\alpha[k]}$  y  $\mathbf{K}_{\beta[k]}$ , respectivamente. Generalmente, se supone que  $p_k \ll p_k$ , por lo que el modelo refleja la situación en la cual, en el canal de mediciones se presentan los ruidos  $\beta[k]$ , considerados típicos, y muy de vez en cuando aparecen interferencias de potencia fuerte  $\alpha[k]$ , que se consideran medidas anómalas.

El vector de valores iniciales  $\mathbf{x}[0]$  de estados del sistema, se asume Gaussiano con valor medio  $\mathbf{m}_{1\mathbf{x}[0]}$  y matriz de covarianza  $\mathbf{R}_{\mathbf{x}[0]}$ . Además, se supone que las sucesiones  $\{\mathbf{w}[k]\}$ ,  $\{\gamma[k]\}$ ,  $\{\alpha[k]\}$  y  $\{\beta[k]\}$  son mutuamente independientes, y a la vez, independientes del vector aleatorio  $\mathbf{x}[0]$ .

La adaptabilidad del algoritmo se basa en el principio de clasificación, como proceso de ordenamiento estadístico, de los resultados de observaciones  $\mathbf{y}[0]$  en dos grupos, tal que sean asociados a los ruidos de la clase  $\{\alpha[k]\}$  ó  $\{\beta[k]\}$ , con la consecuente modificación del algoritmo de filtración, a fin de tener en cuenta el modelo real de las perturbaciones. El procedimiento de clasificación se puede hacer, por ejemplo, empleando el método de toma de decisión bayesiano, mediante el cual la estimación de  $\gamma[k]$  se determina de la siguiente forma,

$$\hat{\gamma}[k] = \begin{cases} 0, & \Lambda_k \{\Delta \mathbf{y}[k]\} \geq \lambda_k \\ 1, & \Lambda_k \{\Delta \mathbf{y}[k]\} < \lambda_k \end{cases}$$

donde  $\Lambda_k \{\Delta \mathbf{y}[k]\}$  es la relación de verosimilitud, que para el caso se determina como

$$\Lambda_k \{\Delta \mathbf{y}[k]\} = \frac{p\{\Delta \mathbf{y}[k] | \gamma[k] = 1\}}{p\{\Delta \mathbf{y}[k] | \gamma[k] = 0\}}$$

siendo  $\lambda_k$  el umbral de decisión. En las anteriores expresiones,  $\Delta \mathbf{y}[k] = \mathbf{y}[k] - \mathbf{C}[k] \mathbf{m}_{1\mathbf{x}[k]}$ , en las cuales el valor medio del proceso  $\mathbf{x}[k]$  se puede determinar en forma recursiva, por ejemplo,  $\mathbf{m}_{1\mathbf{x}[k+1]} = \mathbf{A}[k] \mathbf{m}_{1\mathbf{x}[k]}$ ,  $k \geq 0$ .

La función densidad de probabilidad condicional del vector aleatorio  $\Delta \mathbf{y}[k]$  dado  $\gamma[k]$  está dada por la relación,

$$p\{\Delta \mathbf{y}[k] | \gamma[k]\} = \begin{cases} \sim \mathcal{N}\left(0, \hat{\mathbf{K}}_{\mathbf{x}[k]} - \mathbf{K}_{\alpha[k]}\right), & \gamma[k] = 1 \\ \sim \mathcal{N}\left(0, \hat{\mathbf{K}}_{\mathbf{x}[k]} - \mathbf{K}_{\beta[k]}\right), & \gamma[k] = 0 \end{cases}$$

donde  $\hat{\mathbf{K}}_{\mathbf{x}[k]} = \mathbf{C}[k] \mathbf{K}_{\mathbf{x}[k]} \mathbf{C}^T[k]$ .

De lo anterior, la relación de verosimilitud se calcula como

$$\Lambda_k \{\Delta \mathbf{y}[k]\} = \frac{|\mathbf{K}_{\mathbf{x}[k]+\beta[k]}|^{1/2}}{|\mathbf{K}_{\mathbf{x}[k]+\alpha[k]}|^{1/2}} \exp\left(\frac{1}{2} \Delta \mathbf{y}^T[k] \left(\mathbf{K}_{\mathbf{x}[k]+\beta[k]}^{-1} - \mathbf{K}_{\mathbf{x}[k]+\alpha[k]}^{-1}\right) \Delta \mathbf{y}[k]\right)$$

siendo  $\mathbf{K}_{\mathbf{x}[k]+\eta[k]}^{-1} = \hat{\mathbf{K}}_{\mathbf{x}[k]} + \mathbf{K}_{\eta[k]}$ ,  $\eta \in \{\alpha, \beta\}$ .

La función de covarianza se determina de la siguiente forma recursiva:

$$\hat{\mathbf{K}}_{\mathbf{x}[k+1]} = \mathbf{A}[k] \mathbf{K}_{\mathbf{x}[k]} \mathbf{A}^T[k] + \mathbf{G}[k] \mathbf{K}_{\mathbf{w}[k+1]} \mathbf{G}^T[k]$$

Como resultado, el algoritmo adaptativo para la obtención de solución estable  $\hat{\mathbf{x}}[k]$  en la tarea de filtración, toma la forma,

$$\begin{aligned}\hat{\mathbf{x}}[k] &= \mathbf{A}[k-1] \hat{\mathbf{x}}[k-1] \\ \bar{\mathbf{R}}_{\mathbf{x}}[k] &= \mathbf{A}[k-1] \hat{\mathbf{R}}_{\mathbf{x}}[k-1] \mathbf{A}^T[k-1] + \mathbf{G}[k-1] \mathbf{K}_{\mathbf{w}[k-1]} \mathbf{G}^T[k-1] \\ \tilde{\mathbf{x}}[k] &= \hat{\mathbf{x}}[k] + \mathbf{K}\{\tilde{\gamma}_k, k\} (\mathbf{y}[k] - \mathbf{C}[k] \hat{\mathbf{x}}[k])\end{aligned}$$

La ganancia del filtro adaptativo se determina de la forma ( $\forall k = 1, 2, \dots$ ),

$$\mathbf{K}\{\tilde{\gamma}_k, k\} = \bar{\mathbf{R}}_{\mathbf{x}}[k] \mathbf{C}^T[k] \left( \mathbf{C}[k] \bar{\mathbf{R}}_{\mathbf{x}}[k] \mathbf{C}^T[k] + \tilde{\gamma}[k] \mathbf{K}_{\alpha[k]} + (1 - \tilde{\gamma}[k]) \mathbf{K}_{\beta[k]} \right)^{-1}$$

donde  $\tilde{\gamma}_k = \{\tilde{\gamma}[i] : i = 0, \dots, k\}$ , siendo  $\tilde{\gamma}[i]$  la estimación del correspondiente valor de  $\gamma[i]$ . Los valores iniciales se dan como:  $\hat{\mathbf{x}}[0] = \bar{\mathbf{x}}[0] = \mathbf{m}_{1\mathbf{x}[0]}$ ,  $\bar{\mathbf{R}}_{\mathbf{x}}[0] = \mathbf{R}_{\mathbf{x}[0]}$ .

En esencia, el algoritmo adaptativo de Kalman presentado consiste de dos filtros, los cuales se conectan en concordancia con la decisión tomada sobre el modelo de perturbación presente. Sin embargo, ambos filtros tienen la misma estructura y difieren sólo en el valor de ganancia,  $\mathbf{K}\{\tilde{\gamma}_k, k\}$ , que depende del tipo de perturbación asumida.

El umbral de decisión  $\lambda_k$ , generalmente, se fija de manera experimental para cada aplicación concreta, por ejemplo mediante un modelo de imitación. Aunque existen aproximaciones que intentan fijar este valor de manera automática, así en [50], se presenta un método basado en el análisis discriminante óptimo bayesiano, para una matriz dada de pérdida.

De otra parte, en el algoritmo de Kalman es básica la información que brindan las respectivas funciones de covarianza de los perturbaciones y señales, las cuales no siempre se disponen, por lo tanto, el algoritmo debe tener en cuenta, alguna forma de su estimación. El siguiente algoritmo estima los valores de las funciones de covarianza.

Sea un sistema dinámico lineal en tiempo discreto, descrito por las ecuaciones recurrentes,

$$\begin{cases} \mathbf{x}[k+1] &= \mathbf{A}\mathbf{x}[k] + \mathbf{G}\mathbf{w}[k], \\ \mathbf{y}[k] &= \mathbf{C}\mathbf{x}[k] + \mathbf{v}[k], \quad k = 0, 1, \dots \end{cases} \quad (7.27)$$

donde  $\mathbf{x}[k+1]$  es de estados del sistema con dimensión  $n$ ,  $\mathbf{A}$  es la matriz de transición de orden  $n \times n$ ,  $\mathbf{C}$  y  $\mathbf{G}$  son matrices de orden  $r \times n$  y  $n \times s$ , respectivamente. Se asume que los vectores aleatorios  $\mathbf{w}[k]$  y  $\mathbf{v}[k]$ , con dimensiones respectivas  $s$  y  $r$ , son procesos Gaussianos con correlación nula, y para los cuales se definen los siguientes momentos,

$$\begin{aligned}E\{\mathbf{w}[k]\} &= 0, \quad E\{\mathbf{w}[k] \mathbf{w}[m]^T\} = \mathbf{R}_{\mathbf{w}} \delta_{km} \\ E\{\mathbf{v}[k]\} &= 0, \quad E\{\mathbf{v}[k] \mathbf{v}[m]^T\} = \mathbf{R}_{\mathbf{v}} \delta_{km} \\ E\{\mathbf{v}[k] \mathbf{w}[m]^T\} &= 0\end{aligned}$$

Las matrices de covarianza consideradas,  $\mathbf{R}_{\mathbf{w}}$  y  $\mathbf{R}_{\mathbf{v}}$ , son positivas semidefinidas. Además, se asume que el vector inicial de estados del sistema es Gaussiano con valor medio 0 y matriz de covarianza  $\mathbf{R}_{\mathbf{x}_0}$ . Por último, se asume que el sistema, (7.27), es observable.

De manera similar, al caso del algoritmo adaptativo anterior, la estimación se basa en el análisis de las propiedades de restablecimiento en el tiempo del proceso de filtración, esto es, de la sucesión

$$\Delta \mathbf{y}[k] = \mathbf{y}[k] - \mathbf{C}\hat{\mathbf{x}}[k]$$

El proceso, en su restablecimiento, contiene la nueva información estadística obtenida mediante las observaciones  $\mathbf{y}[k]$ , y la cual sirve para determinar la estimación las respectivas matrices de covarianza  $\mathbf{R}_w$  y  $\mathbf{R}_v$  [11].

Cuando el sistema llega alcanza el estado estable, tanto el coeficiente de ganancia del filtro  $\mathbf{K}$ , como la matriz de covarianza del error extrapolada  $\hat{\mathbf{K}}$  no cambian en el tiempo,

$$\mathbf{K} = \hat{\mathbf{R}}_x \mathbf{C}^\top (\mathbf{C} \hat{\mathbf{R}}_x \mathbf{C}^\top + \mathbf{R}_v)^{-1} \quad (7.28a)$$

$$\hat{\mathbf{R}}_x = \mathbf{A} (\mathbf{I} - \mathbf{K} \mathbf{C}) \hat{\mathbf{R}}_x (\mathbf{I} - \mathbf{K} \mathbf{C})^\top + \mathbf{A} \mathbf{K} \mathbf{R}_v \mathbf{K}^\top \mathbf{A}^\top + \mathbf{G} \mathbf{R}_w \mathbf{G}^\top \quad (7.28b)$$

El par de ecuaciones (7.28a) y (7.28b) sirven para el cálculo de la matriz  $\mathbf{R}_x$ , bien en la estimación óptima, bien en el empleo de algoritmos subóptimos que estiman las matrices de covarianza  $\mathbf{R}_w$  y  $\mathbf{R}_v$ .

Cabe anotar, que la relación estadística entre diferentes valores de observación está dada por su momento conjunto,  $\mathbf{C}_k = E \{ \Delta \mathbf{y}[k] \Delta \mathbf{y}^\top[i - k] \}$ . De otra parte, empleando la definición del proceso de restablecimiento, se pueden obtener las siguientes relaciones importantes en la estimación de las funciones de covarianza buscadas:

$$\begin{aligned} \mathbf{C}_0 &= \mathbf{C} \hat{\mathbf{R}}_x \mathbf{C}^\top + \mathbf{R}_v, \\ \mathbf{C}_k &= \mathbf{C} (\mathbf{A} (\mathbf{I} - \mathbf{K} \mathbf{C}))^{k-1} \mathbf{A} (\hat{\mathbf{R}}_x \mathbf{C}^\top - \mathbf{K} \mathbf{C}_0), \quad k \geq 1 \end{aligned}$$

Básicamente, el algoritmo presente de filtración adaptativa consiste, en que una vez dadas en el primer momento, las estimaciones a priori  $\mathbf{R}_{w[0]}$  y  $\mathbf{R}_{v[0]}$  de las respectivas funciones de covarianza, se realiza el perfeccionamiento de su estimación en el proceso de filtración.

El procedimiento de estimación de las matrices de covarianza es el siguiente:

- Estimación de la matriz  $\mathbf{y}_k$  a partir de la definición

$$\tilde{\mathbf{C}}_k = \frac{1}{N} \sum_{i=k}^N \Delta \mathbf{y}[k] \Delta \mathbf{y}^\top[i - k]$$

siendo  $N - k$  la cantidad de observaciones en análisis.

- Estimación del primer término de multiplicación en (7.28a). Por cuanto,

$$\begin{aligned} \mathbf{C}_1 &= \mathbf{C} \hat{\mathbf{R}}_x \mathbf{C}^\top - \mathbf{C} \mathbf{A} \mathbf{K} \mathbf{C}_0 \\ &\dots \quad \dots \quad \dots \\ \mathbf{C}_n &= \mathbf{C} \mathbf{A}^n \hat{\mathbf{R}}_x \mathbf{C}^\top - \mathbf{C} \mathbf{A} \mathbf{K} \mathbf{C}_{n-1} - \dots - \mathbf{C} \mathbf{A}^n \mathbf{K} \mathbf{C}_0 \end{aligned}$$

entonces

$$\hat{\mathbf{R}}_{\mathbf{x}} \mathbf{C}^T = \mathbf{B} + \begin{bmatrix} \mathbf{C}_1 + \mathbf{CAK}\mathbf{C}_0 \\ \mathbf{C}_2 + \mathbf{CAK}\mathbf{C}_1 + \mathbf{CA}^2\mathbf{K}\mathbf{C}_0 \\ \vdots \\ \mathbf{C}_n + \mathbf{CAK}\mathbf{C}_{n-1} + \cdots + \mathbf{CA}^n\mathbf{K}\mathbf{C}_0 \end{bmatrix}$$

siendo  $\mathbf{B}^+$  la respectiva matriz pseudotranspuesta (de orden  $1 \times n$ )

$$\mathbf{B} = \begin{bmatrix} \mathbf{C} \\ \mathbf{CA} \\ \vdots \\ \mathbf{CA}^{n-1} \end{bmatrix}$$

que corresponde a la multiplicación de las matrices de observabilidad y de transición  $\mathbf{A}$ . Entonces, en calidad de estimación de la matriz  $\mathbf{PC}^T$  se emplea la expresión,

$$\hat{\mathbf{R}}_{\mathbf{x}} \tilde{\mathbf{C}}^T = \mathbf{B}^+ \begin{bmatrix} \tilde{\mathbf{C}}_1 + \mathbf{CAK}\mathbf{C}_0 \\ \tilde{\mathbf{C}}_2 + \mathbf{CAK}\tilde{\mathbf{C}}_1 + \mathbf{CA}^2\mathbf{K}\mathbf{C}_0 \\ \vdots \\ \tilde{\mathbf{C}}_n + \mathbf{CAK}\tilde{\mathbf{C}}_{n-1} + \cdots + \mathbf{CA}^n\mathbf{K}\mathbf{C}_0 \end{bmatrix}$$

- Cálculo de la estimación de la función de covarianza  $\mathbf{R}_{\mathbf{v}}$ , empleando la siguiente expresión  $\tilde{\mathbf{R}}_{\mathbf{v}} = \mathbf{K}\tilde{\mathbf{C}}_0 - \mathbf{C}(\hat{\mathbf{R}}_{\mathbf{x}}\mathbf{C}^T)$
- Cálculo de la estimación de la función de covarianza  $\mathbf{R}_{\mathbf{w}}$ , mediante la solución del siguiente sistema,

$$\begin{aligned} \sum_{i=0}^{k-1} \mathbf{CA}^i \mathbf{G} \tilde{\mathbf{R}}_{\mathbf{w}} \mathbf{G}^T (\mathbf{A}^{i-k})^T \mathbf{C}^T \\ = (\hat{\mathbf{R}}\mathbf{C}^T)^T (\mathbf{A}^{-k}) \mathbf{C}^T - \mathbf{CA}^k \hat{\mathbf{R}}\mathbf{C}^T - \sum_{i=0}^{k-1} \mathbf{CA}^i \tilde{\mathbf{S}} (\mathbf{A}^{i-k})^T \mathbf{C}^T \end{aligned}$$

siendo  $\tilde{\mathbf{S}} = \mathbf{A} \left( -\mathbf{K} (\hat{\mathbf{R}}\mathbf{C}^T)^T - \hat{\mathbf{R}}\mathbf{C}^T \mathbf{K}^T + \mathbf{KC}_0\mathbf{K}^T \right) \mathbf{A}^T$ .

La estimación  $\tilde{\mathbf{C}}_k$ , para de disminuir el coste computacional, se realiza de forma recursiva,

$$\tilde{\mathbf{C}}_k^{N+1} = \tilde{\mathbf{C}}_k^N + \frac{1}{N} \left( \Delta \mathbf{y}[n+1] \Delta \mathbf{y}^T[n+1-k] - \tilde{\mathbf{C}}_k^N \right)$$

donde  $\tilde{\mathbf{C}}_k^l$  es la estimación de  $\mathbf{C}_k$  sobre  $l$  observaciones.



# Bibliografía

- [1] I. I. Privalov, *Vvedenie v teoriju funktsii kompleksnogo peremennogo*. Moskva: Nauka, 1977.
- [2] I. S. Gonorovski, *Circuitos y señales radiotécnicos*. Moscú: Progress, 1972.
- [3] L. A. Zalmanson, *Fourier, Haar, Walsh Transforms and their application in Control, Communications and other Fields*. Moscow: Nauka, 1989.
- [4] L. E. Franks, *Signal Theory*. New Jersey: PH. Englewood Cliffs, 1969.
- [5] R. E. Edwards, *Fourier Series: A Modern Introduction*. Camberra: SV, 1979, vol. I, II.
- [6] G. V. Obrezkov, *Prikladnye Matematicheskie Metody Analiza v Radiotekhnike*. Moskva: VysshajaShkola, 1985.
- [7] N. Wiener, *Nonlinear Problems in Random Theory*, 1st ed. Cambridge, Mass: MIT Press, 1958.
- [8] L. V. Danilov, *Rjady Volterra-Picard v teorii njelinejnyj elektricheskij tsepjey*. Moskva: Radio i Svjaz, 1987.
- [9] V. I. Tijonov, *Nielinijnye prjeobrazovaniya sluchaynyj protsessov*. Moskva: Radio i Svjaz, 1986.
- [10] V. D. Ender and N. A. Verhoeckx, *Discrete-Time Signal Processing. An Introduction*. Cambridge: PHI, 1989.
- [11] V. Bolnokin and P. Chinaev, *Analiz i Sintez Sistem Avtomaticheskogo Upravljenja v EBM*. Moskva: Radio i Svjaz, 1991.
- [12] B. R. Levin, *Teoritieskije osnovy statisticheskoy radiotekhniki*. Moskva: Radio i Svjaz, 1989.
- [13] A. Zaezdnyi, *Osnivy Raschetov po Statisticheskoy Radiotekhnike*. Moskva: Svjaz, 1969.
- [14] G. R. Cooper and C. D. McGuillen, *Probabilistic Methods of Signal and System Analysis*. Pardue University, 1986.
- [15] V. I. Tijonov, *Statisticheskaja radiotekhnika*. Moskva: Radio i Svjaz, 1982.
- [16] A. Lebedev and E. Chernyakovsky, *Verojatnosnije metody v vycheslitel'noy tejnike*. Moskva: Vysshaya Shkola, 1986.
- [17] V. S. Korolyuk and N. Portenko, *Spravochnik por teorii verojatnostey i matematicheskoy statistike*. Moskva: Nauka, 1978.
- [18] A. A. Borovkov, *Estadística Matemática: Estimación, verificación de hipótesis*. Moscú: Mir, 1988.
- [19] E. Lvovskiy, *Statisticheskije metody postrojenija empiricheskij formul*. Moskva: Vysshaja Shkola, 1988.
- [20] W. Lederman, *Handbook of applicable mathematics*, E. Lloyd, Ed. New York: John Wiley Sons, 1984, vol. VI:Statistics.
- [21] S. Yu, *Theory of the Detection and Evaluation of Stochastic Signals (in Russian)*. Moscow: Sovetskoe Radio, 1978.

- [22] P. S. Akimov, A. I. Senin, and V. I. Solenov, *Signaly i ij obrabotkav informatsionnyj sitemaj*. Moskva: Radio i Svjaz, 1994.
- [23] S. Vilenkin, *Statisticheskiye metody issledovaniya sistem avtomaticheskogo regulirovaniya*. Moskva: Sovetskoe Radio, 1967.
- [24] D. Hilbert and R. Kurant, *Métody matematicheskoi fiziki*. Moskva: Gostechizdat, 1951, vol. T.1.
- [25] B. Shajtarin and V. Kovrigin, *Metody spektralnogo otsenivaniya sluchaiynyj proytssessov*. Moskva: Geliyus ARV, 2005.
- [26] J. Bendat and A. Piersol, *Random data: analysis and measurement procedures*. New York, USA: John Wiley and Sons, 1971.
- [27] L. R. Rabiner and B. Gold, *Theory and application of digital signal processing*. Prentice Hall, 1975.
- [28] L. Goldemberg and B. Matiyushkin, *Tsifrovaja Obrabotka signalov*. Moskva: Radio i Svjaz, 1985.
- [29] L. Marple, *Digital Spectral Anaylisis with Applications*. Englewood Cliffs, New Jersey: Prentice Hall, 1987.
- [30] V. I. Tijonov and M. A. Mironov, *Markovskie protseccy*, moskva ed. SovetskoeRadio, 1977.
- [31] R. L. Stratanovich, *Uslovnye Markovskie Protsessy i ij primininje k teorii optimalnogo upravljenja*. Moskva: MGU, 1966.
- [32] Y. I. Paraev, *Vvedenije v statisticheskiju dinamiku protsessov upravljenja i filtratsii*, moskva ed. Sovetskoe Radio, 1976.
- [33] H. Xuedong, A. Acero, and H. Hsiao-wuen, *Spoken Language Processing*. New Jersey: Prentice Hall, 2001.
- [34] A. P. Dempster and N. Laird, "Maximum likelihood from incomplete data via the EM algorithm (with discussion)," *Journal of The Royal Statistica Society B*, vol. 39, pp. 1–38, 1977.
- [35] A. Ljolje, Y. Ephraim, and L. Rabiner, "Estimation of hidden markov models parameters by minimizing empirical error rate," in *Proceedings of International Conference on Acoustics, Speech and Signal Processing*, 1990, pp. 709–712.
- [36] S. E. Levinson, "Continuously variable duration hidden markov models for speech analysis," in *Proceedings of the International Conference in Acoustics, Speech and Signal Processing*, 11, Ed., 1986, pp. 1241–1244.
- [37] L. Rabiner, "A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition," *Proceedings of The IEEE*, vol. 77, no. 2, Febrero 1989.
- [38] M. Do, "Fast approximation of kullback-leibler distance for dependence trees and hidden markov models," *IEEE Signal Processing Letters*, vol. 10, no. 3, pp. 115–118, 2003.
- [39] R. Lyngso and C. Pedersen, "Metrics and similarity measures for hidden Markov models," in *Proceedings of The Seventh International Conference on Intelligent Systems for Molecular Biology*, 1999, pp. 178–186.
- [40] M.Ñechyba and Y. Xu, "Stochastic Similarity for Validating Human Control Strategy Models," *IEEE Transactions On Robotics And Automation*, vol. 14, no. 3, Junio 1998.
- [41] V. I. Tijonov and N. K. Kulman, *Njelinjeynaja filtratsia i kvasikogerentnyi prijem signalov*. Moskva: Sovetskoe Radio, 1975.
- [42] V. I. Tijonov and V.Ñ. Harisov, *Statistichesii Analiz i sintez radiotejnicheskij ustroystv i sistem*. Moskva: Radio i Svjaz, 2004.
- [43] W. B. Davenport and W. L. Root, *An Introduction to the Theory of Random Signals and Noise*. New York: McGraw-Hill, 1958.
- [44] I. S. Guelfand and A. A. Jaglom, *Integrirovanie v Funktsionalnyj Prostranstvaj*. Moskva: YspejMat-Nauk, 1951.

- 
- [45] E. Aved'jan and J. Tzypkin, "Obobschennyi algoritm kachmazha," *Avtomatika i Telemekhanika*, vol. 1, pp. 72–78, 1979.
  - [46] J. Markl, "Uskorenije sjodimosti algoritma kachmazha v sluchaje vremennoi korrelatsii vjednogo protsesssa," *Avtomatika i Telemekhanika*, vol. 8, pp. 70–73, 1980.
  - [47] P. Eikhoff, *System Identification*. Englewoods Cliffs,NJ: Prentice-Hall, 1975.
  - [48] R. Liptzer and A. Shirjaev, *Statistika Sluchainykh Protsessov*. Moskva: Nauka, 1974.
  - [49] M. Ageev, V. Alik, and J. Markov, *Biblioteka Algoritmov 101b-150b*. Moskva: Sovetskoe Radio, 1978.
  - [50] A. Ershov, "Stabil'nye metody otsenki parametrov," *Avtomatika i Telemekhanika*, vol. 8, pp. 66–100, 1978.



# Índice alfabético

- algoritmo
  - máxima esperanza, 206
  - propagación, 204
  - viterbi, 205
- Bayes
  - decisión, 110
  - estimación, 93
  - teorema, 44
- cadena
  - compuesta, 179
  - homogénea, 180
- Chebyshev
  - desigualdad, 47
- coeficiente
  - asimetría, 50
  - exceso, 50
- convolución
  - integral, 10
- correlación
  - estimación, 145
  - función, 69, 117, 124
  - intervalo, 125
  - matriz, 117
  - mutua, 231
- correlograma
  - definición, 161
- costos
  - función, 267
- covarianza
  - función, 69
  - matriz, 77
- decisión
  - óptima, 105
- densidad
  - espectral, 7
- DEP
  - definición, 132
  - estimación, 150
- detección
  - binaria, 268
- dicretización
  - teorema, 25
- Dirichlet
  - condiciones, 8
- distancia, 2
  - Kullback - Leibler, 221
- distribución
  - átomo, 45
- Ecuación
  - aproximación cero, 237
- ecuación
  - Chapman-Kolmogorov, 183
  - Fokker-Plank-Kolmogorov, 197
  - Ito, 194
  - Kolmogorov, 184
  - Kolmogorov-Feller, 196
  - Markov, 188
  - Ricatti, 312
  - Stratanovich, 193, 258
- ensamble, 115
- entropía
  - definición, 80
- ergodicidad
  - definición, 121, 122
- error
  - potencia, 85
- espacio
  - definición, 1
  - probabilístico, 42
- espectro
  - definición, 3
- estacionariedad
  - definición, 116
- estado
  - duración, 217
  - ecuación, 37
  - estable, 19
  - matriz, 37
  - sistema, 31
  - variables, 31, 296
  - vector, 33
- estimación

- definición, 85
- propiedades, 143
- puntual, 86
- seudobayesiana, 113
- FDP
  - definición, 46
  - Fisher, 60
  - Gamma, 58
  - Gauss, 55
  - Poisson, 62
  - Student, 59
  - transformación, 72
- filtración
  - Discreta, 27
- filtro
  - acoplado, 290
  - activo, 293
  - formante, 171
  - pasivo, 293
- Fisher
  - matriz, 278
- Fourier
  - $\mathcal{S}$ -espectro, 5
  - serie, 3
  - transformada, 7
- función
  - base, 2
  - característica, 123
  - generatriz, 68
  - ortogonal, 3
  - probabilidad, 53
  - transferencia, 15, 27
- Hilbert
  - espacio, 49
- hipótesis
  - prueba, 96
- independencia
  - estadística, 43
- información
  - medidas, 78
- intervalo
  - confianza, 91
- Ito
  - integral, 194
- Jensen
  - desigualdad, 209
- K-L
  - transformada, 129
- Kachmazh
  - método, 303
- Kalman
  - filtración, 308
- Kalman-Bucy
  - filtro, 312
- Kolmogorov
  - axiomas, 42
- Kotelnikov
  - teorema, 25
- Laplace
  - integral, 58
  - transformada, 10
- mínimos cuadrados
  - modelo, 99
- MacLaurin
  - serie, 51
- Markov
  - cadena, 178
  - modelos ocultos, 201
  - procesos, 127, 177
  - sucesiones, 198
  - transformación, 242, 249
- Minimax
  - detección, 271
  - Neymann-Pearson, 271
- modelo
  - AR, 172
  - ARMA, 172
  - MA, 172
  - paramétrico, 171
- momentos
  - centralizados, 47
  - cumulantes, 51
  - estimación, 144
  - iniciales, 46
  - relación, 52
- Neyman-Pearson
  - criterio, 97
- norma, 2
- operador
  - lineal, 12, 14, 234
  - método, 16
- pérdida
  - función, 105
- periodograma
  - algoritmo, 164
  - Bartlett, 158
  - Blackman-Tukey, 162
  - definición, 152
  - método, 157
  - Welch, 160
- promedios
  - tiempo, 116
- RBG
  - definición, 136
- redundancia
  - definición, 82
- regresión
  - estimación, 153
  - lineal, 100
  - modelo, 98
- respuesta

- a impulso, 15, 16, 27
- riesgo
  - función, 107
- señal
  - aleatoria, 115
- spline
  - definición, 99
- Stieltjes
  - integral, 141
- transformación
  - lineal, 12
- transformada
  - discreta, 26
- trayectoria, 115
- valores
  - medios, 46
- varianza
  - propiedades, 48
- ventana
  - espectral, 166
- verosimilitud
  - función, 88
- Volterra
  - series, 21, 243
- Wiener
  - filtración, 307
  - método, 20
- Wiener-Hopf
  - ecuación, 155, 284
- Wiener-Jinchin
  - transformada, 134, 232
- Yule-Walker
  - ecuaciones, 174

Apoyo Académico



Ingeniería